

The Association of Coal-Fired Power Plants and the Mortality of COVID-19

Alexander Irwin¹ and Robert Hodgson[#]

¹McDowell High School, Erie, PA, USA

[#]Advisor

ABSTRACT

With the outbreak and spread of severe acute respiratory syndrome coronavirus 2 (COVID-19), the world was ushered into a global pandemic. One of the most alarming aspects of the disease is the high mortality rate. Previous research has shown that long term exposure to air pollution has been correlated with the mortality rate of respiratory diseases. The focus of this study was to determine if there was any association between the mortality of COVID-19 and one of the primary producers of air pollution, coal-fired power plants. Using data from the US Energy Information Administration, John Hopkins University, and census department, a chi-square test and a poisson regression analysis were conducted to determine if living in proximity to coal-fired power plants had any effects on the mortality of COVID-19. The chi-square test results showed that there was no statistical significance as the variables showed independence. These results illustrate that there is no association between coal-fired power plants and the mortality rate of COVID-19. To expand on the results of the chi-square test, a poisson regression analysis was performed to account for the presence of confounding variables. This analysis showed similar results to the chi-square test, but due to issues with outliers in the data causing overdispersion, the model was unable to be accurately conducted, making all results inconclusive. With the inconclusive results of the poisson regression analysis, the conclusions drawn from the chi-square test were not able to be generalized as they were not verified in the presence of confounding variables.

Introduction

A new world health emergency has emerged due to the spread of the severe acute respiratory syndrome coronavirus 2 (SARS-Cov2 or COVID-19). This virus originated from Huanan Seafood Wholesale Market in Wuhan, located within China's Hubei Province (Wu 2020). The disease was originally carried by bats and was spread to humans through an unknown intermediary animal (Singhal 2019, Morens 2020). On January 1st, 2020, the Huanan market was closed down after the Chinese government notified the World Health Organization (WHO) about the outbreak (Wu 2020). The number of cases increased exponentially as the disease quickly spread to other provinces (Wu 2020). The spread of the disease was exacerbated when people who traveled to Wuhan for the Chinese New Year spread the disease to other nations. On January 21st, 2020, the first case was confirmed in the United States and from that point the case numbers increased dramatically (Hauck 2020). There are currently 32,778,906 cases in the United States and 582,837 deaths as of May 11, 2021 (World Health Organization 2021).

Both symptomatic and asymptomatic people may transmit the disease, as it can travel up to six feet. Infection can occur when large droplets are transmitted through touching the facial area or inhaling the droplets (Singhal 2019). The virus's standard side effects include fever, cough, sore throat, headache, fatigue, myalgia, and breathlessness. The disease can quickly progress into pneumonia, respiratory failure, and ultimately can lead to death (Singhal 2019).

Literature Review

Underlying Conditions in Patients

Certain underlying conditions can affect the progression and severity of the disease. In March of 2020, a study was performed by COVID-NET to determine the hospitalization rates and characteristics of patients hospitalized with COVID-19. They found that 90% of hospitalized patients had one or more underlying conditions previous to COVID-19 (Shikha 2020). The most common underlying conditions were obesity, chronic lung disease, diabetes mellitus, hypertension, and cardiovascular disease (Shikha 2020). People with chronic lung diseases are at high risk for COVID-19 because of impaired lung function (Gallay 2020). Researchers investigated over 5,000 patients with COVID-19 that were omitted to hospitals to see how pre-existing respiratory illnesses affect mortality (Signes-Costa 2020). Their results showed that patients with lung disease had a 10% increase in COVID-19 mortality. This established that there is a relationship between COVID-19 mortality and having pre-existing lung diseases (Signes-Costa 2020).

Air Pollution

Respiratory diseases, such as COVID-19, are often impacted by air pollution (Wu 2020). Air pollution refers to substances that were released into the air which have harmful or poisonous effects. These substances include carbon dioxide (CO₂), nitrogen dioxide (NO₂), sulfur dioxide (SO₂), and particulate matter. The substances can cause breathing issues and impact how your body combats respiratory diseases (Doiron 2018). Previous research conducted a study where they examined COVID-19 death counts from over 3000 counties up to April 22, 2020 (Wu 2020). They found an increase of one micrometer/cubic meter of air pollution resulted in an eight percent increase in COVID-19 mortality (Wu 2020). This established that COVID-19 mortality is impacted by increased amounts of air pollution.

Coal-fired power plants use the combustion of coal to generate electricity. The burning of coal releases air pollution into the atmosphere (Energy Education 2020). This pollution has been shown to cause adverse health effects on humans and can lead to respiratory issues (Energy Education 2020). Because the highest amount of pollution settles down in the immediate area around the power plant, the surrounding communities will experience the most severe health effects. These effects could include asthma, Chronic Obstructive Pulmonary Disease (COPD), or possible death (Kravchenko 2018). Previous research showed that air pollution from coal-fired power plants shortened the lives of over 30,000 Americans each year (Schneider 2000). Therefore, coal-fired power plants could potentially impact mortality in reference to certain respiratory diseases. Further research found that the 1918 influenza pandemic was expedited by coal-fired power plants' usage (Clay 2018). Researchers discovered that cities that experienced an increase in coal usage resulted in tens of thousands of excess deaths in 1918 relative to cities that used less coal with similar baseline health and pre-pandemic socioeconomic conditions (Clay 2018).

Research Gap

Previous research has made the connection between COVID-19 mortality and air pollution (Wu 2020). Studies have also discovered a link between coal-fired power plants and the mortality of respiratory diseases, but there has not yet been any research on how the mortality rate of COVID-19 is affected by the presence of coal-fired power plants. This leads to the primary focus of this research which will look at the association between coal-fired power plants and COVID-19 mortality. This generated the research question: Is there an association between living in proximity to coal-fired power plants and the mortality of COVID-19 in the United States? In regard to this question it was hypothesized that there will be association between coal-fired power plants and the mortality rate of COVID-19. This was made due to the finding of previous researchers which found that coal-fired power plants were associated with higher levels of mortality of the 1918 influenza pandemic (Clay 2018).

Methodology

COVID-19

To analyze COVID-19 mortality, county level mortality data was collected. Specifically, data was collected on the number of total cases and the number of deaths for each county. This information was used to calculate the death rate for each county by dividing the deaths by the total cases. The decision to analyze mortality was made in accordance with prior research proving that mortality is an effective way to measure a disease's severity (Wu 2020, Clay 2018). Research has also used county-level information to verify that air pollution affects the mortality of COVID-19 (Wu 2020). County-level information was used in this study as it was the smallest data set currently available to the general public, thus providing the most accurate and narrow content available for research. The information and data on the mortality of COVID-19 was gathered by using information from John Hopkins University (Wu 2020). This database is updated daily by state officials and provides the most accurate and current information. It also provided the number of cases and deaths for each county. This data was used to determine the mortality rate for each county in my research. The mortality of 307 counties was analyzed in this study. This represents the total number of counties with coal-fired power plants and reference counties, which is furthered in discussion about coal-fired power plants. The data for the counties that were used in this study were from January 21, 2020, through February 4, 2021. The first case of COVID-19 was reported on January 21, and the data was collected on February 4. This total included the mortality of all of the counties with coal-fired power plants as well as the reference counties.

Coal-Fired Power Plants

The location of every coal-fired power plant in each county was another variable analyzed in this study. The US Energy Information Administration (US EIA) provided a map that gave detailed information on the location of every currently operating coal-fired power plant in the United States. Previous research has used the US EIA to determine the locations of fuel-fired power plants (Carpenter 2012). From this source, the location of the 262 currently operating coal-fired power plants in the United States was found. In addition, the location of counties without coal-fired power plants was also collected. Previous research has compared the hospitalization rates for zip codes with fuel-fired power plants to "clean" zip codes or zip codes without a power plant that were not directly affected by fuel-fired power plants (Carpenter 2012). This same principle was applied in this study to counties and mortality rates to act as a reference or control group to compare the mortality data from counties with coal-fired power plants. This data was from counties that were not exposed to pollution from coal-fired power plants. From John Hopkins University, the number of cases and deaths for 45 reference counties was collected. This number was chosen as there were 45 states with currently active coal-fired power plants. The five states not included are Connecticut, Maine, Massachusetts, Rhode Island, and Vermont due to their lack of coal-fired power plants. There was one reference county chosen from each state in order to get a wide variety of counties as well as to include all areas of study. For each county, information on mortality rates as well as confounding variables was also collected.

Confounding Variables

The confounding variables that were addressed include population density and median annual income. Population density was chosen because diseases can spread more quickly in areas with a higher population density (Carpenter 2012, Wu 2020). Median household income was chosen to control for the relative socioeconomic class for each county. Socioeconomic class is important as poor people are much more vulnerable to poor health than rich people (Carpenter 2012). The information regarding the confounding variables was gathered from the U.S. Census Department to quantify these variables (Wu 2020, Carpenter 2012). The Census Department provided accurate information on population density and median household income. The information gathered from the Census Department is from 2019, so some variables are subject to natural change; however, the changes would not be so significant to impact the

results of the study (Carpenter 2012). Therefore, the information was utilized in this study. The population density and median household income for 307 counties were used in this study.

Chi-Square Test

For the first half of the analysis, a Pearson chi-square test was performed. A chi-square test is used to determine if there is a relationship between two variables. Research has used a log-linear analysis to analyze a similar topic of study (Carpenter 2012). A chi-square test is a simplified version of a log-linear analysis as you only analyze two variables, while in a log-linear analysis you can analyze multiple variables. For this study, the two variables studied were COVID-19 mortality and coal-fired power plants. The variables have to be categorical and since the mortality data is numerical, the variables were modified for how it is represented in the test. For the covid-19 mortality variable, it was found that the average mortality rate of COVID-19 for the United States is 1.8%. In order to categorize the data, it was determined if the mortality rates for each county are higher or lower than the national average and the categories labeled as such. The coal-fired power plant data was divided into two categories: coal or no coal. These represent counties with coal-fired power plants (coal) and the reference counties without coal-fired power plants (no coal). If there is an association between coal-fired power plants and a higher mortality rate then the original hypothesis would have been proven correct.

Before the chi-square test can begin, a null hypothesis was observed. The null hypothesis assumes that the two variables are independent of each other. If the hypothesis is confirmed, then the two variables are independent, but if the hypothesis is refuted, then there is an association between the two variables.

Poisson Regression Analysis

A Poisson regression analysis was conducted in order to study the mortality and coal-fired power plant variables, as well as to include the confounding variables of population density and median household income. Poisson regression is used to analyze a set of data with numerical and categorical variables. Poisson regression is often compared to log-linear regression as they are very similar in how the dependent variables are categorized. The difference arises in what variables can be used. In a log-linear analysis, all of the variables have to be categorical, while in Poisson regression analysis the independent variable is numerical data and the dependent variables are categorical. For this analysis, four variables were selected, which are COVID-19 mortality, coal-fired power plants, median household income, and population density. The COVID-19 mortality was the independent variable, while the others were all dependent. The mortality variable was measured by using the raw number of deaths for each county gathered from John Hopkins University. The coal-fired power plants were divided into the same categories as the chi-square test: coal and no coal. The population density was classified as either rural or urban, with rural having a population of fewer than 1000 people per square mile and urban having more than 1000 people per square mile. The median household income was divided into four quartiles based on their ranges. This study used the average median household income from the entire US as the median value. The maximum and minimum values were chosen based on the states included in the analysis. The first and third quartiles were found by averaging the minimum and maximum with the median. There were 256 counties included in the analysis. This number was smaller than the number used in the chi-square test as some counties were not included because they did not have a median household income that fits between the ranges of median household income mentioned above.

Results

Chi Square Test Analysis

Table 1 shows a summary of all of the cases used in the analysis. There were 307 counties that were analyzed in the chi-square test. All of the counties used were valid and met the requirements needed to be inputted into the test. When a chi-square test is performed, a null hypothesis must be drawn. The null hypothesis portrays no association between the variables studied or that they are independent of each other. The null hypothesis for this study is that there is no association between coal-fired power plants and COVID-19 mortality.

TABLE 1.

Case Processing Summary					
Cases					
Valid					
Missing					
Total					
	N	Percent	N	Percent	Percent
Counties * Mortality	307	100.0%	0	0.0%	307

As shown in Table 2, there are two categories for the type of county, counties with coal-fired power plants and counties without coal-fired power plants. These are represented by labels of coal or no coal. The first row of cells for each category is the observed count, labeled as “count” in the table, and is the frequency of the number of counties that have a higher or lower mortality rate than was observed during the test. The total number of counties observed with coal-fired power plants was 262 counties. Within this total, there were 110 counties with higher mortality rates than the national average and 152 counties with lower mortality rates than the national average. The total number of counties in this sample without coal-fired power plants was 45. Among these, 21 counties had a higher mortality rate than the national average, and 24 had a lower mortality rate than the national average.

The second row of each category in the table is the expected count. This is the predicted frequency under the assumption that the null hypothesis is true. The total number of expected counties with coal-fired power plants is the same (262) as the observed count. In this total, 111.8 counties were expected to be higher than the national average and 150.2 counties were expected to be lower than the national average. The total number of counties without coal-fired power plants is also the same as the observed count (45). In regards to this total, 19.2 were expected to be higher than the national average and 25.8 were expected to be lower than the national average.

The purpose of the crosstabs table is to give a general idea of the association between the two variables and to compare the observed and expected counts. If the observed values diverge substantially from the expected counts then there will be a higher chi-square score. This means that there is a significant relationship between the variables and the null hypothesis will be rejected. The observed and expected counts in this study did not diverge significantly. This means that it is assumed that there is not a strong association between the two variables, but the chi-square statistic will definitively exemplify whether there is any association.

TABLE 2.

Counties * Mortality Crosstabulation					
			Mortality		Total
			higher	lower	
Counties	Coal	Count	110	152	262
		Expected Count	111.8	150.2	262.0
	No Coal	Count	21	24	45
		Expected Count	19.2	25.8	45.0
Total		Count	131	176	307
		Expected Count	131.0	176.0	307.0

In Table 3, the two values of importance to this study are the chi-square score and the p-value. The value of the chi-square score is .344. This is a relatively low chi-square value and reveals that there is relatively little association between the two values. The p-value is used to determine if the null hypothesis can be confirmed or rejected. This is represented in the first row under the asymptotic significance cell. There is a p-value of .557, which is relatively high and shows that the results are not significant by confirming the null hypothesis. This means the two variables are independent of each other. In order to have significance, the p-value should be lower than the designated alpha value, which for this test is .05.

TABLE 3.

Chi-Square Tests					
	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.344	1	.557		
Continuity Correction ^a	.179	1	.672		
Likelihood Ratio	.342	1	.559		
Fisher's Exact Test				.625	.334
N of Valid Cases	307				

a. 0 cells (.0%) have an expected count less than 5. The minimum expected count is 19.20.
b. Computed only for a 2x2 table

The results of this test show that there is no association between counties with coal-fired power plants and higher mortality rates. This was in contrast with the findings of previous research as they have shown correlation between coal-fired power plants and mortality (Carpenter 2012, Clay 2018). In an effort to rectify the disparity between my results and those of previous researchers a Poisson regression model was conducted. This model was used to analyze the association between counties with coal-fired power plants and the number of deaths in each county (another measure of mortality). The test included confounding variables that could potentially influence the number of deaths in each county, which were the median household income and population density.

Poisson Regression Analysis

Table 4 was used to confirm the dependent variable: the number of deaths. It also explained that the probability distribution is Poisson and the link function is the natural logarithm. This table confirmed that the test being run is Poisson regression and that the natural log was used to analyze the data.

TABLE 4.

Model Information	
Dependent Variable	Number of Deaths
Probability Distribution	Poisson
Link Function	Log

Table 5 illustrated the number of cases included in this analysis as 256 counties. All of these values were complete and included in the analysis. There were no values that were excluded from the statistical analysis.

TABLE 5.

Case Processing Summary		
	N	Percent
Included	256	100.0%
Excluded	0	0.0%
Total	256	100.0%

Table 6 looked at the number and percentage of cases in each category for each independent variable. The first independent variable is counties. This was divided into two groups: coal and no coal. There were 224 counties with coal-fired power plants, which represented 87.5% of the total. There were 32 counties without coal-fired power plants, which represented 12.5% of the total. The second independent variable is population density, represented by rural or urban labels. There were 238 counties that had a rural population, which represented 93.0% of the total. There were 18 counties with an urban population, which represented 7.0% of the total. The third variable is median household income, which is divided into four categories. These represent the four income quartiles in the United States and are labeled as quartiles 1-4 (shown in the table as Q1-Q4). There were 98 counties in the first quartile, which represents 38.3% of the total. There were 87 counties in the second quartile and they represented 34% of the total. There were 50 counties in the third quartile, which represented 19.5% of the total. There were 21 counties in the fourth quartile and this represented 8.2% of the total. These numbers were not well distributed between the categories for each variable due to the relative frequencies of each category. This creates groups that are highly unbalanced, which can cause issues with model fit. This could potentially mean that the data does not fit well with the poisson regression model.

TABLE 6.

Categorical Variable Information				
			N	Percent
Factor	Counties	Coal	224	87.5%
		No Coal	32	12.5%
		Total	256	100.0%
	Population Density	rural	238	93.0%
		urban	18	7.0%
		Total	256	100.0%
	Income	Q1	98	38.3%
		Q2	87	34.0%
		Q3	50	19.5%
		Q4	21	8.2%
		Total	256	100.0%

Table 7 checked the data for problems and determined if there was overdispersion in the analysis. This was done by comparing the ratio of the variance to the mean. The variance was calculated by squaring the standard deviation and dividing that value by the mean (dependent variable). The value of the ratio in this study is 2039.221808. The ratio should be close to 1, which means that the mean and variance are equal. The value in this study is very large which means that there was a lot of overdispersion. This was due to the very high values of mortality in some of the cases. These values are outliers and have a significant effect on the results. From this, one can assume that the data does not fit with the model, but this was further determined in the goodness of fit table.

TABLE 7.

Continuous Variable Information						
		N	Minimum	Maximum	Mean	Std. Deviation
Dependent Variable	Number of Deaths	256	0	8926	246.19	708.545

Table 8 provided many measures to assess how well the model fit the data. To determine how well the model fits with the data, the value/df of the Pearson chi-square row was used. The value in this study is 319.648, which is a very high value. If there is equidispersion, then this value should be 1.00. This value is a lot larger than 1.00, which again means that there was a lot of overdispersion. This would violate the fifth assumption in order to perform a Poisson regression analysis. The fifth assumption states that the variance should be equal to the mean, but in this study, the variance is a lot larger than the mean. There was a large sample size (256 cases) and the model allows samples with a high case number to have a larger amount of dispersion, so this could possibly offset the violation. However, there was a significant amount of overdispersion, which cannot be negated by the large sample size.

TABLE 8.

Goodness of Fit			
	Value	df	Value/df
Deviance	64269.722	250	257.079
Scaled Deviance	64269.722	250	
Pearson Chi-Square	79911.969	250	319.648
Scaled Pearson Chi-Square	79911.969	250	
Log Likelihood ^a	-32911.597		
Akaike's Information Criterion (AIC)	65835.195		
Finite Sample Corrected AIC (AICC)	65835.532		
Bayesian Information Criterion (BIC)	65856.466		
Consistent AIC (CAIC)	65862.466		

Dependent Variable: Number of Deaths
Model: (Intercept), Counties, Population Density, Income
a. Information criteria are in smaller-is-better form.
b. The full log-likelihood function is displayed and used in computing information criteria.

The Omnibus Test (Table 9) was utilized to see if all of the independent variables collectively improve the model, in comparison to the intercept-only model, which has no independent variables added. The measure taken from this table is the p-value which is represented by the significance column. The p-value is 0, which means that the model is not statistically significant.

TABLE 9.

Omnibus Test		
Likelihood Ratio Chi-Square	df	Sig.
90851.451	5	.000

Dependent Variable: Number of Deaths

Model: (Intercept), Counties, Population Density, Income
a. Compares the fitted model against the intercept-only model.

Table 10 was used to show the significance of each individual independent variable, which is shown in the significance column. From this table, the p-values were extracted for each variable. The counties, population density, and median household income variables had a p-value of 0, thus meaning that they were statistically insignificant. All of the p-values were 0, which means that all of the independent variables were not statistically significant.

TABLE 10.

Tests of Model Effects			
Source	Type III		
	Wald Chi-Square	df	Sig.
(Intercept)	426008.032	1	.000
Counties	30.331	1	.000
Population Density	94840.006	1	.000
Income	12377.094	3	.000

Dependent Variable: Number of Deaths
Model: (Intercept), Counties, Rural or Urban, Income Quartiles

Table 11 provided the coefficient estimates (*B column*) and the exponentiated values of the coefficients (*ExB column*). The ExB column is usually more informative than the former. For example, there is a value of 1.099 for the counties with coal-fired power plants. This means that there is a 9.9% increase in the number of deaths for countries with coal-fired power plants. However, this can not be conclusive or proven because the p-value for this category is 0. This means that the independent variable or number of deaths is insignificant. This is the same for all of the other variables of counties, population density, and median household income. The p-values are all 0, which means that there is no significance attributed to any of the variables.

TABLE 11.

Parameter Estimates					
Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test
			Lower	Upper	Wald Chi-Square
(Intercept)	6.397	.0275	6.343	6.451	54168.425
[Counties=Coal]	.095	.0172	.061	.128	30.331
[Counties=No Coal]	0 ^a
[Population Density=rural]	-2.597	.0084	-2.614	-2.581	94840.006
[Population Density=urban]	0 ^a
[Income=Q1]	1.240	.0272	1.187	1.294	2080.139
[Income=Q2]	.464	.0275	.410	.518	284.962
[Income=Q3]	1.515	.0273	1.462	1.569	3082.908
[Income=Q4]	0 ^a
(Scale)	1 ^b				

Parameter Estimates					
Parameter	Hypothesis Test				
	df	Sig.	ExB	Lower	Upper
(Intercept)	1	.000	600.170	568.693	633.389
[Counties=Coal]	1	.000	1.099	1.063	1.137
[Counties=No Coal]	.	.	1	.	.
[Population Density=rural]	1	.000	.074	.073	.076
[Population Density=urban]	.	.	1	.	.
[Income=Q1]	1	.000	3.457	3.277	3.646
[Income=Q2]	1	.000	1.591	1.507	1.679
[Income=Q3]	1	.000	4.551	4.314	4.802
[Income=Q4]	.	.	1	.	.
(Scale)					

Dependent Variable: Number of Deaths
Model: (Intercept), Counties, Rural or Urban, Income Quartiles
a. Set to zero because this parameter is redundant.
b. Fixed at the displayed value.

Discussion

The purpose of this study was to determine if living in proximity to a coal-fired power plant was associated with COVID-19 mortality. The hypothesis was formulated to be that living in proximity to a coal-fired power plant was associated with COVID-19 mortality. This hypothesis was tested by performing a chi-square test and Poisson regression analysis.

The chi-square test found that there was no statistical significance in the variables used because the p-value was relatively high and confirmed the null hypothesis, which hypothesized that the variables were independent of each other. The confirmation of the null hypothesis showed that there was no relationship between COVID-19 mortality and living in proximity to a coal-fired power plant.

In order to expand on the conclusions reached in the chi-square test, a Poisson regression analysis was performed. This analysis showed that the results were not statistically significant. This is shown by the p-value of the model which was zero, which suggests that the model is statistically insignificant. Every independent variable was also shown to be insignificant as they also had p-values of zero. However, the model was not able to be accurately conducted due to problems with the data. The first of these problems was the number of outliers included in the data shown by the continuous variable information table. This caused the standard deviation to be a lot higher than the mean of the number of deaths. It also causes a lot of overdispersion, which means that the model might not fit with the data. This high amount of overdispersion violated the fifth assumption of a Poisson regression analysis, which states that there should not be much overdispersion in the model. A violation of any assumption means that the conclusions generated are circumstantial and need to be investigated further.

Limitations

A possible limitation of this study is the data on mortality. County-level mortality data was used as that is the only data available to the general population. Prior research has previously used zip code level data to prove the correlation between fuel-fired power plants and respiratory disease mortality, but due to lack of availability zip code level mortality was not able to be evaluated (David Carpenter). This could impact my results as coal-fired power plants could have a greater effect on the mortality of the surrounding zip code than on the entire county.

The size of the county could also pose limitations. Some of the counties have a very large area, while others have a very small area. This could potentially pose issues as the smaller counties might be more impacted by the pollution from the coal-fired power plants and the larger counties could be less impacted by the pollution because they have more space to be dispersed.

The quantity of counties represented could limit the conclusions drawn. There were 45 counties used as a reference to compare the data on counties with coal-fired power plants due to issues with time and data availability. In order to get the full scope and most accurate mortality information, there needs to be a larger number of counties analyzed. Also, the conclusions drawn from this data could potentially not be able to be generalized to the rest of the population as the counties analyzed in this study only comprise a small portion of the population. The results could possibly not be the same for other parts of the country.

There also could possibly be limitations on the coal-fired power plant data. During the data collection process, it was discovered that some counties have more than one coal-fired power plant within the borders of the county. This could influence the results as this could have a greater effect on the mortality rate of these counties as there is more coal being produced and more pollution being released into the air. In addition, some of the pollution from the coal-fired power plants could spread into other counties. This could affect the mortality rate of those counties even though they were not part of the original analysis of this study.

The Poisson regression analysis looked at how median household income and population density affect mortality. There are other variables that could also impact the mortality and number of deaths in a county. These factors include racial and age percentages and the number of hospital beds and ventilators. The former could affect mortality by showing how the virus disproportionately affects one group of people or one specific age range. The latter explains how the lack of or adequacy of medical equipment can affect mortality.

Future Research

Future researchers looking to expand on the conclusions drawn by this study should remove the outliers from the Poisson regression analysis. This would remove the overdispersion and negate the violation of the fifth assumption. It could also impact the significance of the variables, which would ultimately allow definite conclusions to be drawn. These conclusions could support or refute the original hypothesis that there is association between COVID-19 mortality and coal-fired power plants.

Investigating other factors that could influence the mortality rate is an important extension of this analysis. For instance, analyzing other sources of pollution, in addition to coal-fired power plants. This could include cities that use more automotive transportation, or other generating power sources, such as oil or natural gas. Previous research has shown that air pollution is associated with coal-fired power plants, so it is important to investigate all of the possible sources of air pollution.

Variables that could impact the mortality rate of the group studied should also be considered. The variables analyzed in this study were limited to median household income and population density. Other confounding variables could include racial and age percentages, as well as, the availability of medical technology. From the racial or age percentages, future research can determine whether a specific group of people are more likely to die from this disease. The availability of medical technology, such as hospital beds and ventilators, can determine whether a community is adequately prepared to fight this virus or a similar one and if additional government funding is needed.

Conclusion

Before the analysis was conducted, the hypothesis was that there would be association between coal-fired power plants and COVID-19 mortality. This hypothesis was refuted through the results of the chi-square test which illustrated statistically insignificant results. Through analysis of these results, this study was able to conclude that there was no association between coal-fired power plants and COVID-19 mortality. However, with the Poisson regression analysis having inconclusive results, the conclusions from the chi-square test cannot be generalized as they were not able to be verified in the presence of confounding variables.

Acknowledgments

I would like to thank Mr. Hodgson, my AP Research teacher, for all of his encouragement and guidance throughout the entire process of creating and developing my research paper. He never gave up on me, even when I had severe doubts about my abilities. It was through his teaching, as well as his constant and unwavering support that I was able to achieve such a high level of academic writing. I cannot express my deepest gratitude for everything that he has done for me.

References

American Journal of Respiratory and Critical Care Medicine,
www.atsjournals.org/doi/citedby/10.1164/rccm.202007-2638LE.

Brook RD, Franklin B, Cascio W, et al. Air pollution and cardiovascular disease: a statement for healthcare professionals from the Expert Panel on Population and Prevention Science of the American Heart Association. *Circulation* 2004;109(21):2655- 71. doi: 10.1161/01.CIR.0000128587.30041.C8 [published Online First: 2004/06/03]

Bureau, US Census. Census.gov, www.census.gov/.

By. "Interpreting Chi Square Results in SPSS." EZ SPSS Tutorials, 2 Dec. 2019, ezspss.com/interpreting-chi-square-results-in-spss/.

CDC. "Coronavirus Disease 2019 (COVID-19)." Centers for Disease Control and Prevention, 11 Feb. 2020, www.cdc.gov/coronavirus/2019-ncov/covid-data/covid-net/purpose-methods.html.

Ciencewicki J, Jaspers I. Air pollution and respiratory viral infection. *Inhal Toxicol* 2007;19(14):1135-46. doi: 10.1080/08958370701665434 [published Online First: 2007/11/08]

Clay K, Lewis J, Severnini E. Pollution, infectious disease, and mortality: evidence from the 1918 Spanish influenza pandemic. *J Econ Hist* 2018;78(4):1179-209. doi: 10.1017/S002205071800058X [published Online First: 2018/10/02]

Coticini E, Frediani B, Caro D. Can atmospheric pollution be considered a cofactor in the extremely high level of SARS-CoV-2 lethality in Northern Italy? *Environ Pollut*

Contini, D., & Costabile, F. (2020, April 13). Does Air Pollution Influence COVID-19 Outbreaks? Retrieved September 08, 2020, from <https://www.mdpi.com/2073-4433/11/4/377/htm>

“COVID-19 United States Cases by County.” Johns Hopkins Coronavirus Resource Center, coronavirus.jhu.edu/us-map.

[Crossref], [PubMed], [Web of Science ®], [Google Scholar]
<https://ajph.aphapublications.org/doi/abs/10.2105/AJPH.79.5.623>

Cui Y, Zhang ZF, Froines J, et al. Air pollution and case fatality of SARS in the People's Republic of China: an ecologic study. *Environ Health* 2003;2(1):15. doi: 10.1186/1476-069X-2-15 [published Online First: 2003/11/25] 42. Pope CA, 3rd. Respiratory disease associated with community air

Desjardins, Jeff. “Mapped: Every Power Plant in the United States.” *Visual Capitalist*, 17 June 2020, www.visualcapitalist.com/mapped-every-power-plant-in-the-united-states/.

Doiron D;de Hoogh K;Probst-Hensch N;Fortier I;Cai Y;De Matteis S;Hansell AL; “Air Pollution, Lung Function and COPD: Results from the Population-Based UK Biobank Study.” *The European Respiratory Journal*, U.S. National Library of Medicine, pubmed.ncbi.nlm.nih.gov/31285306/.

Dominici F, Peng RD, Bell ML, et al. Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *JAMA* 2006;295(10):1127-34. doi: 10.1001/jama.295.10.1127 [published Online First: 2006/03/09]

Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 2020 doi:10.1016/S1473-3099(20)30120-1 [published Online First: 2020/02/23]

Eric, Coker, et al. The Effects of Air Pollution on COVID-19 Related Mortality in Northern Italy. *Fondazione Eni Enrico Mattei (FEEM)*, 2020, www.jstor.org/stable/resrep25159. Accessed 26 Aug. 2020.

Graziano Onder, M. (2020, May 12). COVID-19 Case-Fatality Rate and Characteristics of Patients Dying in Italy. Retrieved September 22, 2020, from <https://jamanetwork.com/journals/jama/fullarticle/2763667>

Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 - COVID-NET, 14 States, March 1–30, 2020. (2020, April 16). Retrieved September 23, 2020, from <https://www.cdc.gov/mmwr/volumes/69/wr/mm6915e3.htm?s+cid=mm6915e3+w>

“Item 1000/2490: Repositorio Nacional.” Item 1000/2490 | Repositorio Nacional, covid-19.conacyt.mx/jspui/handle/1000/2490.

Jaakkola J. J. K., Paunio M., Virtanen M., Heinonen O. P. Low-level air pollution and upper respiratory infections in children. *Am. J. Public Health* 1991; 81: 1060–1063 [Google Scholar]

J.M.K.C. Donev et al. (2019). Energy Education - Coal fired power plant [Online]. Available: https://energyeducation.ca/encyclopedia/Coal_fired_power_plant. [Accessed: May 11, 2021].

Kravchenko J, Lyerly HK; The impact of coal-powered electrical plants and coal ash impoundments on the health of residential communities. *N C Med J*. 2018;79(5):289-300 (in this issue).

Morens, D. M., Breman, J. G., Calisher, C. H., Doherty, P. C., Hahn, B. H., Keusch, G. T., Kramer, L. D., LeDuc, J. W., Monath, T. P., & Taubenberger, J. K. (2020). The Origin of COVID-19 and Why It Matters, *The American Journal of Tropical Medicine and Hygiene*, 103(3), 955-959. Retrieved Apr 15, 2021, from <https://www.ajtmh.org/view/journals/tpmd/103/3/article-p955.xml>

- “Mortality Analysis.” Johns Hopkins Coronavirus Resource Center, coronavirus.jhu.edu/data/mortality.
- “Myalgia.”, 2019, www.hopkinsmedicine.org/health/conditions-and-diseases/myalgia.
- “Particulate Matter (PM2.5, PM10-2.5, and PM10) and Children's Hospital Admissions for Asthma and Respiratory Diseases: A Bidirectional Case-Crossover Study.” Taylor & Francis, www.tandfonline.com/doi/abs/10.1080/15287390801907459.
- “Pennsylvania COVID-19 Numbers.” Department of Health, www.health.pa.gov/topics/disease/coronavirus/Pages/Cases.aspx.
- Philip K. Hopke, Daniel Croft, Wangjian Zhang, Shao Lin, Mauro Masiol, Stefania Squizzato, Sally W. Thurston, Edwin van Wijngaarden, Mark J. Utell, David Q. Rich, Changes in the acute response of respiratory diseases to PM2.5 in New York State from 2005 to 2016, *Science of The Total Environment*, Volume 677, 2019, Pages 328-339, ISSN 0048-9697, <https://doi.org/10.1016/j.scitotenv.2019.04.357>.
- PMC, E. (n.d.). Retrieved September 24, 2020, from <https://europepmc.org/article/med/32026671>
- “Poisson Regression Analysis Using SPSS Statistics.” How to Perform a Poisson Regression Analysis in SPSS Statistics | Laerd Statistics, statistics.laerd.com/spss-tutorials/poisson-regression-using-spss-statistics.php.
- Pope C. A., III. Respiratory disease associated with community air pollution and a steel mill, Utah valley. *Am. J. Public Health* 1989; 79: 623–628
- Provisional Death Counts for Coronavirus Disease 2019 (COVID-19). (2020, September 23). Retrieved September 23, 2020, from <https://www.cdc.gov/nchs/nvss/vsrr/COVID19/>
- Schneider, Conrad G. Death, Disease, and Dirty Power. Mortality and health damage due to air pollution from power plants. United States: N. p., 2000. Web.
- Shahzad Baig K, Yousaf M (2017) Coal Fired Power Plants: Emission Problems and Controlling Techniques. *J Earth Sci Clim Change* 8: 404. doi: 10.4172/2157-7617.1000404
- Signes-Costa J, Núñez-Gil IJ, Soriano JB, Arroyo-Espliguero R, Eid CM, Romero R, Uribarri A, Fernández-Rozas I, Aguado MG, Becerra-Muñoz VM, Huang J, Pepe M, Cerrato E, Raposeiras S, Gonzalez A, Franco-Leone F, Wang L, Alfonso E, Ugo F, García-Prieto JF, Feltes G, Abumayyaleh M, Espejo-Paeres C, Jativa J, Masjuan AL, Macaya C, Carbonell Asíns JA, Estrada V; HOPE COVID-19 investigators. Prevalence and 30-Day Mortality in Hospitalized Patients With Covid-19 and Prior Lung Diseases. *Arch Bronconeumol*. 2021 Apr;57 Suppl 2:13-20. doi: 10.1016/j.arbres.2020.11.012. Epub 2020 Dec 16. PMID: 33423874; PMCID: PMC7744014.
- Singhal, T. A Review of Coronavirus Disease-2019 (COVID-19). *Indian J Pediatr* 87, 281–286 (2020). <https://doi.org/10.1007/s12098-020-03263-6>
- “U.S. Energy Information Administration - EIA - Independent Statistics and Analysis.” United States - Maps - U.S. Energy Information Administration (EIA), www.eia.gov/state/maps.php.
- Wu, Y., Chen, C., & Chan, Y. (2020, March). The outbreak of COVID-19: An overview. Retrieved September 08, 2020, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7153464/>

Xiao Wu, Rachel C. Nethery, Benjamin M. Sabbath, Danielle Braun, Francesca Dominici medRxiv
2020.0405.20054502; dio: <https://dio.org/10.1101/2020.04.05.20054502>

Xiaopeng Liu, Lawrence Lessner, and David O. Carpenter 2012; Association between Residential Proximity to Fuel-Fired Power Plants and Hospitalization Rate for Respiratory Diseases; Environmental Health Perspectives 120:6 CID: <https://doi.org/10.1289/ehp.1104146>

Xu Z, Shi L, Wang Y, et al. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *Lancet Respir Med* 2020;8(4):420-22. doi: 10.1016/S2213- 2600(20)30076-X [published Online First: 2020/02/23]

Zhu, Y., Xie, J., Huang, F., & Cao, L. (2020, April 15). Association between short-term exposure to air pollution and COVID-19 infection: Evidence from China. Retrieved September 08, 2020, from <https://www.sciencedirect.com/science/article/pii/S004896972032221X>