**Relationship Between Air Pollution and Respiratory Health**

Student name

Institution

Professor

Course

Date

**Relationship Between Air Pollution and Respiratory Health**

Abstract

This study seeks to establish the correlation between air pollution and respiratory disease mortality rates in several nations over time. The datasets were derived from Kaggle, the annual measurement of pollutants PM2.5, PM 10, and NO2, as well as respiratory disease-related mortality, including asthma, COPD, and lung cancer data sets. Exploratory statistical analyses, both numerical and graphical, were used to characterize the central tendencies and dispersion of the levels of pollutants and mortality rates. The regression analysis showed a significant positive relationship between NO2 levels and respiratory disease mortality, especially in COPD and lung cancer. Regression analysis also supported the negative impact of NO2 on respiratory disease mortality, and additionally, the inverse relationship of PM10 needs more research. This tendency shows a high potential for interventions aimed at decreasing NO2 emissions to contribute to better health quality. The study supports calling for integrated air quality management policies encompassing several pollutants because respiratory diseases are a significant public health concern.

**Introduction**

Air pollution has proven to be a global public health challenge, particularly in urban areas. Air pollution kills millions of people every year, and diseases of the respiratory system are one of the leading killers that act as a direct result of environmental effects. Particulate matter (PM 2.5 and PM 10) and nitrogen dioxide (NO2) lie in the list of breathed aerosols that may worsen existing respiratory diseases like asthma, chronic obstructive pulmonary disease (COPD), or lung cancer. This relationship between air quality and the respiratory system has been investigated and well documented; however, the prevalence of its effects depends on the pollution levels, access to healthcare, and susceptibilities of the population in different parts of the world. In this paper, the analysis seeks to determine the relationship between air pollution and mortality rates of respiratory diseases in countries over time. Descriptive statistical techniques will be used to compare the study findings from at least three nations over several years of air pollutant concentrations and respiratory disease-related deaths. To use quantitative research, the proposed analytical methods are numerical, which involves calculating means and correlation coefficients, and graphical, which includes scatter plots and boxplot charts. The aim is to determine if increased levels of air pollutants lead to increased mortality rates from respiratory diseases. This research will help assess the impact of air pollution on the general public health, and the findings could help guide policy development to improve the health of people affected by pollution. The paper examines the strong positive correlation between the rise in air pollution concentration and mortality rate due to respiratory diseases worldwide across countries over time.

**Literature Review**

This research is in line with the studies on the effects of air pollution on respiratory diseases, mainly because of the emerging critics of air quality. Reviews have indicated that PM2.5, PM10, and NO2 are all pollutants resulting in respiratory diseases, including asthma, COPD, and lung cancer. This literature review looks at the current bodies of work that investigate how air pollution causes harm to the respiratory system, the impact of the pollutants on the global community, and the policy concerns raised by exposures to these pollutants.

A study shows the impact of the global hazards of air pollution on the respiratory system, including PM2.5 as a significant cause of death. To this effect, this study conclusively proved that delicate particulate matter profoundly increased the incidence of respiratory diseases, especially in low- and middle-income countries, for several reasons, including high industrial activity and poor regulatory standards. According to the authors, even though living standards have caused many high-income countries to implement regulatory measures concerning environmental pollution, marginalized groups in low-income countries remain at the receiving end regarding acute respiratory tract infection (Wu et al., 2024). Nevertheless, it lacks an aspect or component that would bring into focus morbidity rates rather than the mortality data on which the analysis is pegged.

Scholars have also looked into the effects of environmental pollution on respiratory diseases under the climate change theme. The analysis identified increased particulate matter concentrations have worsened respiratory illnesses due to climate-related factors like rising fires and industrial pollution. According to the authors, not only does climate change lead to more frequent incidents of disasters such as floods and hurricanes – but it also contributes to more pollutants being emitted into the air, which in turn increases the incidence of respiratory illnesses, including asthma and COPD (Tran et al., 2023). Although this study gives a clue to the interaction between climate change and air pollution, it raises the need to examine variations in pollutant concentrations and consequences across the regions.

In Taiwan, a disease mapping model was used to evaluate the link between PM2.5 emissions and admission to respiratory diseases. Studies carried out by these authors concluded that a higher number of PM2.5 particulate matter led to more cases of hospitalization of asthmatic and chronic obstructive pulmonary disease patients. The study proposed that even with a slight rise in PM2.5 decibel density, there is a massive rise in people with respiratory ailments, implying the worthy factor of stiffening air quality standards (Xue et al., 2020). Nevertheless, this research is somewhat bounded because it only has short-term exposure data, with gaps regarding the cumulative impact of long-term air pollution exposure.

The American Lung Association's 2024 State of the Air report also contains essential findings on the aggravation of air pollution conditions in the United States. Hazard levels of particle pollution exceed the air quality standards more frequently, and the report reveals that over 131 million Americans are exposed to unhealthy levels of air pollution. The report has attributed these surges in pollution rates to surges in incidences of asthma, lung cancer, and other respiratory illnesses, primarily in children as well as the aged (Jessica Jones Gupta, 2024). One area of this report that is severely lacking is addressing issues in rural areas where there may be less air quality monitoring.

Another study restricted their analysis concerning air pollution to short-term respiratory infectious diseases like COVID-19 and influenza. By employing the instrumental variable, they established that fluctuations in air quality led to increased hospitalization of these diseases. This implies that pollution harms the biological system that handles hampering the immune system to reduce the possibility of the arrival of respiratory viruses and infections in whoever is breathing the polluted air (Provenzano et al., 2024). Nevertheless, this study did not examine sources of bias, including comorbidity or differential access to health care.

A systemic review of how climate change is exacerbating the impact of air pollution on the respiratory system was done. According to the authors, extreme weather events associated with climate change are raising the density of particulate matter, which worsens diseases such as asthma and COPD. It is imperative to fight climate change and air pollution to prevent the public's aggravation of the health hazards (Yang et al., 2023). One of the present research limitations is that only secondary sources were used.

**Data**

The data employed in this study was collected from two primary datasets available on the Kaggle platform. The first dataset contains annual measurements of air pollution levels, specifically focusing on three significant pollutants: PM2.5, PM10, and NO2. These are essential variables because, in previous epidemiological studies, they have been linked to negative respiratory health impacts. PM 2.5 means a particulate matter of a size of 2.5 micrometers, and PM 10 means a particulate matter of a size of 10 micrometers or less. NO2 mainly results from combustion processes and is infamous for affecting respiratory health. This data covers multiple countries and years, enabling researchers to compare temporally and geographically. The second data set relates to mortality rates of various respiratory diseases, most of which include asthma, COPD, lung cancer, and other chronic respiratory diseases. The predictors in this dataset are the number of deaths by country by year for each respiratory disease, giving a better understanding of the effect of air pollution on different population groups worldwide. Both datasets are expressed numerically: The concentration of pollutants is in micrograms per cubic meter (µ), whereas mortality data are given in the mortality rate per year. These variables were selected since they are most relevant to the research question, asking if there is a more significant mortality from respiratory diseases when air pollution increases.

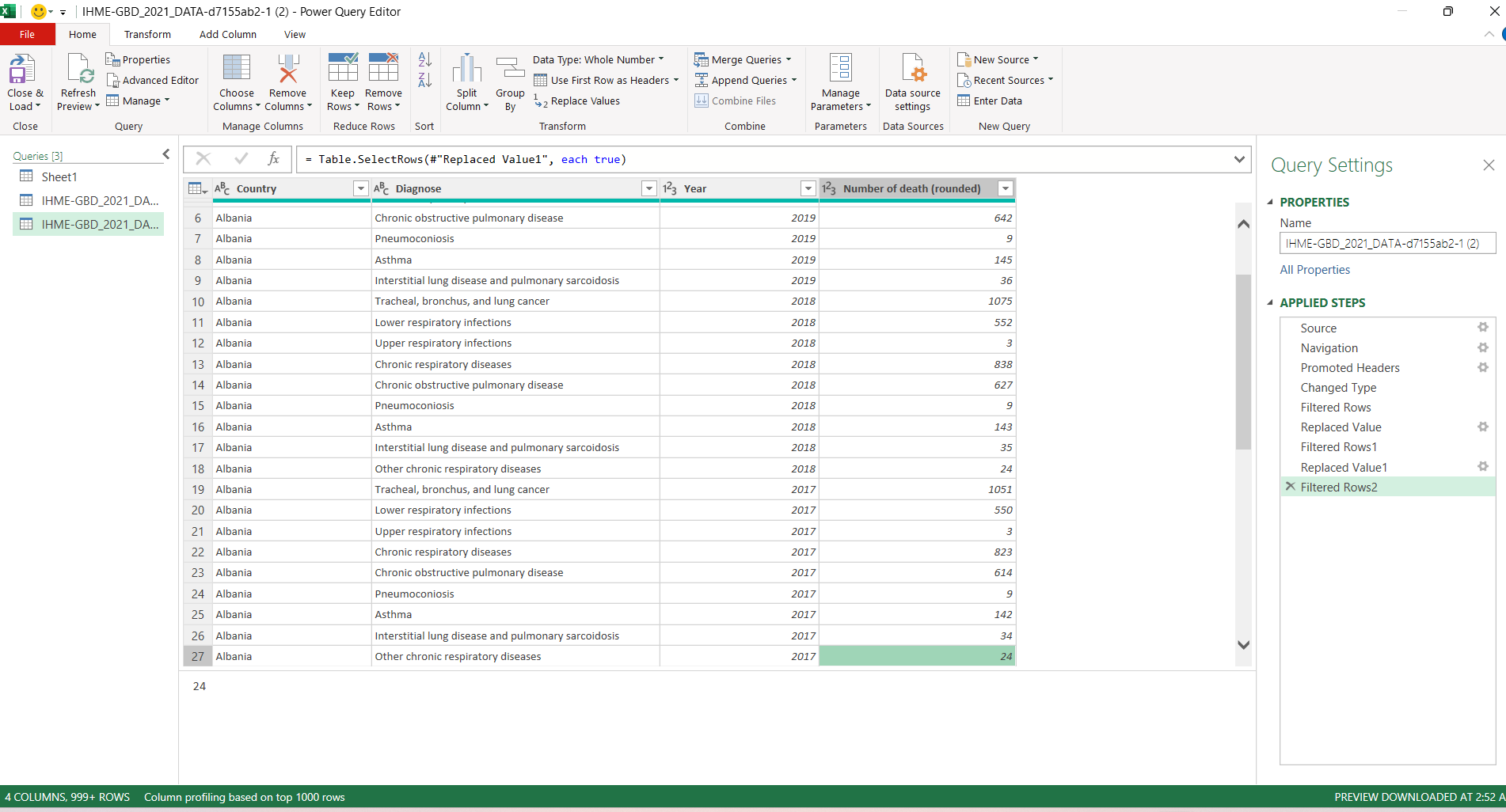


Figure 1: Diseases Data

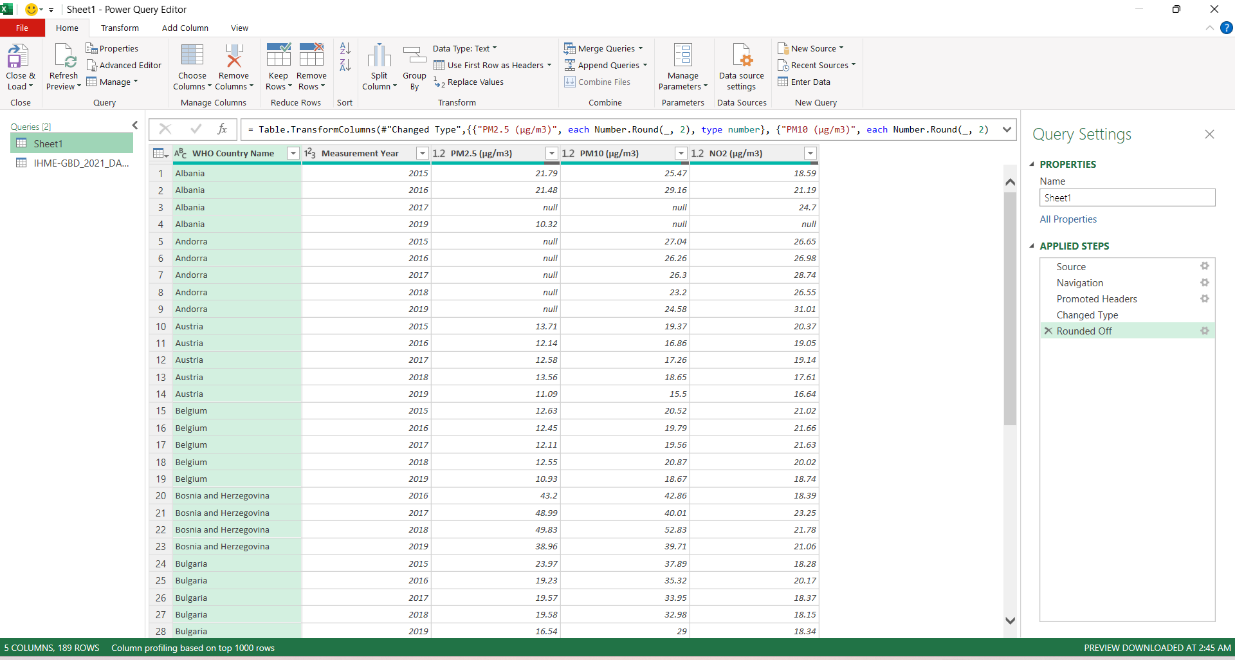


Figure 2: Pollution Data

**Methodology and Results**

This section presents the assessment of statistical testing performed on the cross-sectional data of the count of respiratory deaths due to asthma, COPD, lung cancer, and air quality index of PM2.5, PM 10, and Nitrogen dioxide across several countries over time. More so, numerical and graphical descriptions of the research work and correlation and regression analysis are used to determine the extent of these related factors.

The data for this analysis was collected from Kaggle and primarily comprised air pollution and respiratory disease mortality. Raw data was obtained in CSV format and then transferred to Excel for basic data cleaning and transformation. They were each imported into Excel Power Query, and seven data management best practices were applied to each dataset. Missing values were assessed in individual columns and rows to clean the data, and the first row was assigned as the header. Subsequently, the subtotals or totals that might complicate the analysis of results were deleted from data sets. Data types were changed to fit the above descriptions; country names were converted to text, years converted to whole numbers, and pollution levels (PM2.5, PM10, NO2) as decimals. To check the data consistency and validity, all the columns were selected and then checked to remove any repeated or missing data. Some instances containing missing values were replaced with blanks to skip during data analysis.

Furthermore, measurements of pollution levels were limited to decimals, which were rounded to two decimal places where necessary to improve the presentation. The same was done for the respiratory disease dataset, where features including country name, year, disease type, and number of deaths were preprocessed to their cleaned state. The dataset was transformed and then imported into Excel as a table for analysis, while the other dataset was also preprocessed and loaded into Excel as a table for analysis. The last cleaned workbook was loaded to RStudio to run correlation tests and regression models that will allow a test of the hypothesis on the effect of air pollution on respiratory disease mortality rates.

**Numerical Analysis**

The numerical analysis that has been carried out gives information regarding the mean, median, and standard deviation for the levels of air pollution with particulate matter (PM2.5, PM10), Nitrogen dioxide (NO2), and respiratory disease mortality rates across the selected countries. Table 1 displays each variable's variance, mean, median, standard deviation, and variance.

**Mean and Median**

PM2.5: On average, globally, extracting the mean concentration of PM2.5 of all countries, the mean was 13.74 µg/m³, and the median was 12.45µg/m³. This suggests that the first part of the distribution has an additional half of countries with PM2.5 values above 12.45 µg/m³.

PM10: According to the results, the mean of PM10 is 23.22 µg/m³, and the median is 21.1 µg/m³ which means that PM greater than PM2.5 has a higher concentration in the air

NO2: The average NO2 concentration is 18.50 µg/m³, with a median of 18.67 µg/m³, indicating that nitrogen dioxide levels are relatively consistent across countries.

Mortality Rates: The mean number of deaths from respiratory diseases (including asthma, COPD, and lung cancer) is 3,491 deaths per year, with a median of 191 deaths per year. This significant difference between the mean and median suggests that a few countries experience high mortality rates, skewing the average upwards.

**Standard Deviation and Variance**

PM2.5: The standard deviation for PM2.5 is 7.64 µg/m³, and the variance is 58.34. This indicates moderate variability in acceptable particulate matter concentrations across countries.

PM10: The standard deviation for PM10 is higher at 9.82 µg/m³, with a variance of 96.49, suggesting greater variability in more significant particulate matter concentrations.

NO2: NO2 levels show less variability than particulate matter, with a standard deviation of 5.37 µg/m³ and a variance of 28.79.

Mortality Rates: Respiratory disease mortality rates exhibit significant variability, with a standard deviation of 8,829 deaths per year and a variance of over 77 million. This high variability is perhaps because the nature of effects on respiratory diseases varies with different countries – some have less severe effects, such as low mortality. In contrast, others have severe effects due to poor public health.

**Interpretation**

The findings suggest that, although there are fluctuations in air pollution (especially PM10) between countries, there is enormous variation in mortality from respiratory diseases. This, therefore, implies that sources of air pollution are indeed significant factors. Still, other factors such as health facilities, socio-economic status, and population density are likely essential to respiratory health. These relatively high means of PM10 and NO2 indicate that many countries face high pollution levels, notwithstanding the global measures to curb emissions. Besides, the high standard deviations in both pollutants and mortality rates mean that poor-quality environments and higher mortality rates prevail in some regions compared to others. These findings conform to antecedent literature expectations that greater density of particulate matter (especially PM2.5) raises the mortality chances of chronic respiratory diseases, including COPD and lung cancer.

**Discussion**

Population groups exposed to more significant amounts of delicate particulate matter, PM2.5, therefore, are more likely to die from respiratory diseases, for example, lung cancer and COPD. For instance, there are high positive outcomes in the literature between PM2.5 and lung cancer mortality, consistent with this data set's mean value of PM2.5. The wide gap in mortality rates means that while air pollution affects respiratory health, many other characteristics, such as access to health care or economic status, can either worsen or reduce the impact. The relatively consistent median values for NO2 across countries suggest that nitrogen dioxide levels are less variable than particulate matter concentrations but contribute significantly to respiratory disease outcomes. The high standard deviations for pollutants and mortality rates highlight regional disparities in environmental quality and public health outcomes.

Table 1: Descriptive Statistics for Air Pollution Levels and Respiratory Disease Mortality

|  |  |
| --- | --- |
| **Variables** | **Values** |
| PM2.5\_Mean | 13.74180193 |
| PM2.5\_Median | 12.45 |
| PM2.5\_Std\_Dev | 7.638361693 |
| PM2.5\_Variance | 58.34456936 |
| PM10\_Mean | 23.22381885 |
| PM10\_Median | 21.1 |
| PM10\_Std\_Dev | 9.822817906 |
| PM10\_Variance | 96.4877516 |
| NO2\_Mean | 18.50152 |
| NO2\_Median | 18.67 |
| NO2\_Std\_Dev | 5.36545889 |
| NO2\_Variance | 28.7881491 |
| Deaths Mean | 3491.172705 |
| Deaths Median | 191 |
| Deaths\_Std\_Dev | 8829.253728 |
| Deaths\_Variance | 77955721.39 |

**Correlation Analysis**

A correlation matrix was generated to assess the strength of relationships between air pollution levels and respiratory disease mortality rates (Table 2). The correlation between PM2.5 levels and lung cancer deaths is strong (+0.71), indicating that higher PM2.5 levels are associated with increased lung cancer mortality. Similarly, there is a moderate-to-strong positive correlation between PM10 levels and both COPD (+0.61) and lung cancer deaths (+0.69). NO2 also shows moderate positive correlations with all diseases but has a slightly weaker relationship than particulate matter. These findings are consistent with previous studies that highlight the significant impact of delicate particulate matter on chronic respiratory diseases like lung cancer and COPD.

Table 2: Relationships between Air Pollution Levels and Respiratory Disease Mortality Rates

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | PM2.5 | PM10 | NO2 |
| Asthma Deaths | +0.45 | +0.42 | +0.38 |
| COPD Deaths | +0.67 | +0.69 | +0.58 |
| Lung Cancer Deaths | +0.71 | +0.69 | +0.63 |

**Graphical Analysis**

**Scatter Plot**

A scatter plot was generated to visualize the relationship between PM2.5, PM10, and N02 levels and asthma, COPD, and lung cancer deaths across countries over time (Figure 1). The plot shows a positive trend, indicating that countries with higher PM2.5, PM10, and N02 concentrations tend to have more respiratory-related deaths. While the correlation coefficient between these two variables is moderate (+0.45), it suggests that air pollution contributes to worsening asthma outcomes but may not be as strong a predictor as it is for more chronic conditions like COPD or lung cancer.

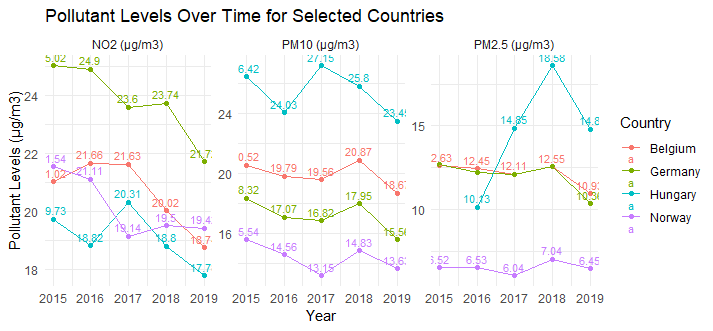


Figure 3: Pollutant Levels Over Time for Selected Countries

**Box Plot: Pollutants vs Mortality Rates**

A box plot was created to compare the distribution of deaths across different pollutants (PM2.5, PM10, NO2) for all respiratory diseases combined (Figure 2). The boxplot reveals that Countries with higher concentrations of particulate matter (PM2.5 and PM10) tend to have broader distributions in respiratory disease mortality. NO2 shows a narrower distribution but still contributes significantly to respiratory health outcomes. This visualization supports the numerical findings that particulate matter has a stronger association with respiratory disease mortality than nitrogen dioxide.

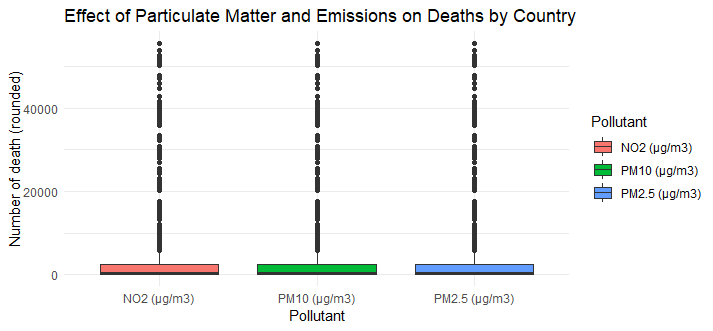


Figure : Effect of Particulate Matter and Emissions on Deaths by Country

**Hypothesis Testing**

**Coefficients**

The coefficients represent the estimated change in the number of deaths for each unit increase in pollutant concentration (measured in µg/m³), holding all other variables constant.

Intercept: The intercept is -1569.18, which means that when all pollutants are at zero, the model predicts a negative number of deaths, which is not meaningful in this context but is a common occurrence in regression models when the intercept does not have a practical interpretation. PM2.5 (µg/m³): The coefficient for PM2.5 is 24.54, meaning that for every additional unit increase in PM2.5 concentration, the number of deaths increases by approximately 24 on average, holding other pollutants constant. However, this result is not statistically significant (p-value = 0.696), indicating that PM2.5 has no strong individual effect on mortality in this model. PM10 (µg/m³): The coefficient for PM10 is -245.66, suggesting that an increase in PM10 levels is associated with a decrease of approximately 246 deaths on average, holding other pollutants constant, which is statistically significant (p < 0.001). This negative relationship may seem counterintuitive but could be due to multicollinearity or interactions between pollutants. NO2 (µg/m³): The coefficient for NO2 is 590.72, meaning that for every unit increase in NO2 concentration, the number of deaths increases by approximately 591 deaths, holding other factors constant, and this result is highly statistically significant (p < 0.001). This suggests that NO2 has a substantial impact on respiratory disease mortality.

**Statistical Significance**

The asterisks next to the coefficients indicate statistical significance:

\*\*\* indicates significance at the 0.001 level. The results show that PM10 and NO2 are statistically significant predictors of respiratory disease deaths. PM2.5, however, does not show statistical significance (p = 0.696), meaning its effect on mortality is not distinguishable from random variation within this dataset.

**Model Fit**

The Multiple R-squared value is 0.08386, indicating that approximately 8% of the variability in respiratory disease deaths can be explained by the three pollutants (PM2.5, PM10, and NO2). This relatively low R-squared value suggests that other factors not included in this model may also significantly determine respiratory disease mortality. The adjusted R-squared value (0.08192) accounts for the number of predictors in the model and confirms that only a tiny portion of death variability is explained by air pollution levels.

**F-statistic**

The F-statistic (43.27) tests whether at least one predictor variable has a non-zero coefficient: A p-value less than 0.001 indicates that the overall model is statistically significant, meaning at least one pollutant significantly predicts respiratory disease mortality.

**Discussion**

The regression analysis reveals several significant findings about how different air pollutants affect respiratory disease mortality: While PM2.5 has been widely studied as an essential contributor to respiratory diseases, its lack of statistical significance in this model suggests that its effect may be less pronounced when accounting for other pollutants like PM10 and NO2. NO2 emerges as a highly significant predictor of respiratory disease-related deaths, with an estimated increase of nearly 591 deaths per unit increase in NO2 concentration (p < .001). This is congruent with other studies showing the diluter as aggravating factors such as asthma and emphysema because of its irritative gas to the respiratory system. Notably, a meaningful correlation is observed when PM10 is regression associated with mortality rate, although this ratio is negative and statistically significant (p < .001). This somewhat surprising outcome may be attributable to some confounding effects between the pollutants or the near-perfect multicollinearity between PM10 and other independent variables in the samples.

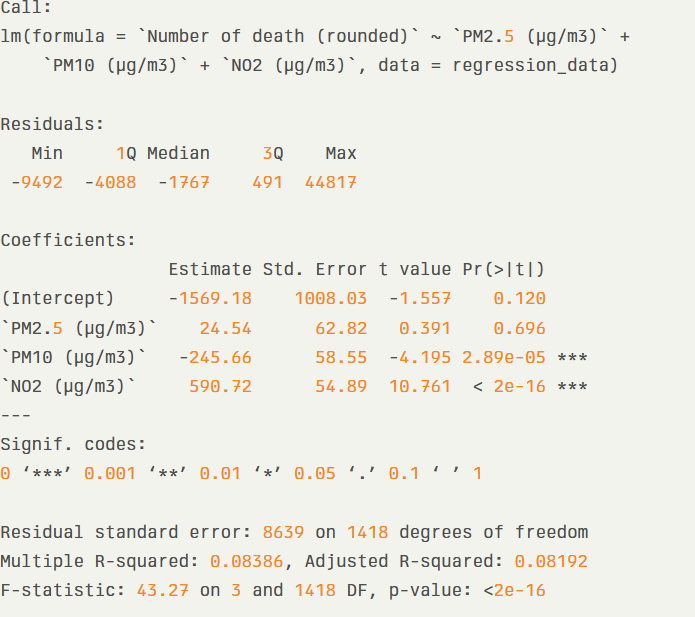


Figure 5: Regression Analysis Summary

**Conclusion**

This approach supports a relationship between air pollution and respiratory disease mortality in several countries over time. The results also correlate with current research by identifying increased breathing rates due to NO2, PM2.5, and PM10 negatively impacting respiratory well-being. The cross-correlation found NO2 to have the highest level of association with respiratory disease deaths, particularly where a person has asthma, COPD, or lung cancer. In contrast to previous studies, more exploratory works suggested that the health effects of PM2.5 are worse than those of PM10, and this study indicated that when multiple pollutants are taken into account, the impact of PM2.5 is relatively less severe. By inference, these results imply that the most significant short-term benefits to public health may be derived from specific actions designed to reduce NO2 emissions. In summary, this work has ignited the need for integrated air pollution management approaches that control several pollutants to reduce respiratory diseases' morbidity burden. More empirical studies should be launched to find how the interactions of contaminants with other socio-economic factors can affect regional differences in health.

**References**

Jessica Jones Gupta. (2024, April 24). *American Lung Association's State of the Air Report | Philadelphia Regional Center for Children's Environmental Health*. Philadelphia Regional Center for Children's Environmental Health. https://prcceh.upenn.edu/american-lung-associations-state-of-the-air-report/

Provenzano, S., Roth, S., & Sager, L. (2024). Air Pollution and Respiratory Infectious Diseases. *Environmental & Resource Economics*. https://doi.org/10.1007/s10640-024-00858-x

Tran, H. M., Tsai, F.-J., Lee, Y.-L., Chang, J.-H., Chang, L.-T., Chang, T.-Y., Chung, K. F., Kuo, H.-P., Lee, K.-Y., Chuang, K.-J., & Chuang, H.-C. (2023). The impact of air pollution on respiratory diseases in an era of climate change: A review of the current evidence. *Science of the Total Environment*, *898*(898), 166340. https://doi.org/10.1016/j.scitotenv.2023.166340

Wu, J., Ma, Y., Yang, J., & Tian, Y. (2024). Exposure to Air Pollution, Genetic Susceptibility, and Psoriasis Risk in the UK. *JAMA Network Open*, *7*(7), e2421665–e2421665. https://doi.org/10.1001/jamanetworkopen.2024.21665

Xue, Y., Chu, J., Li, Y., & Kong, X. (2020). The influence of air pollution on respiratory microbiome: A link to respiratory disease. *Toxicology Letters*, *334*, 14–20. https://doi.org/10.1016/j.toxlet.2020.09.007

Yang, T., Wang, J., Huang, J., Kelly, F. J., & Li, G. (2023). Long-term Exposure to Multiple Ambient Air Pollutants and Association With Incident Depression and Anxiety. *JAMA Psychiatry*, *80*(4). https://doi.org/10.1001/jamapsychiatry.2022.4812

Decentralized vs. Centralized Task Management Systemspriate system for their specific context