A regional study to evaluate the impact of coal-fired power plants on Lung cancer Incident Rates

Dinesh Pal Mudaranthakam (dmudaranthakam@kumc.edu)
University of Kansas Medical Center
https://orcid.org/0000-0001-9767-1158

Sam Pepper
University of Kansas Medical Center

Alexander Alsup
University of Kansas Medical Center

Byron Gajewski
The University of Kansas Medical Center

Keywords:

Posted Date: November 19th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2047023/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Lung cancer is currently the leading cause of cancer death worldwide due to its high incidence rate and low survival rate (1). Despite its high mortality, early screening of lung cancer is underemphasized in public campaigns compared to other cancers (2). Many risk factors contribute to lung cancer, with the predominant cause being the inhalation of toxic chemicals which includes tobacco smoke and industrial pollution (3). The combustive process of coal power production releases 84 different compounds that are designated as hazardous air pollutants by the United States Environmental Protection Agency (4). These compounds can cause several diseases in both humans and animals, as demonstrated by national-level research studies conducted in Southeast Asia (5). As of 2020 coal-powered power plants contributed to 34% of the overall power generation across Kansas (6). One ton of coal only generates 2,460 kWh of electricity whereas Wyandotte county alone requires 2,300 kWh (7; 8).

With that context, our goal was to assess how the coal-fired power plants across the state of Kansas are related to lung cancer incidence in their surrounding area. We found that areas within the immediate vicinity of two coal-power plants had higher incidence rates of lung cancer compared to areas with no coal-power plants. Additionally, modeling lung cancer incidence based on vicinity to plants with covariates revealed a significant relationship between poverty, age, and lung cancer incidence. Individuals living in poverty are predisposed to healthcare-related bankruptcy and cost-associated treatment nonadherence (9). They are also shown to smoke more which is a known risk factor for lung cancer (10). Limiting affordable housing for these individuals to areas containing significant risk factors for lung cancer is irresponsible and potentially exploitative. Further studies on this topic should examine additional socioeconomic and lung cancer risk factors as well.

Introduction:

The process of cell division within the human body is necessary for the replacement of old and damaged cells (11). When this cell division occurs as intended, the human body can replenish damaged tissue and continue to operate efficiently. However, cell division can become pathological and result in uncontrolled cell proliferation, which is referred to as cancer. Cancer can occur in any part of the human body and is a deadly and notoriously difficult disease to treat (12).

Among various cancer types, lung cancer is very aggressive and hard to treat (13). Lung cancer is considered a leading cause of death across the globe among both men and women. The survival rate for lung cancer patients is very low compared to other cancer types, with the main reason being the late diagnosis of lung cancer compared to other cancers such as breast cancer (2). Screening methodologies have improved in recent years, but it is still difficult to implement proper screening guidelines for lung cancer (14). Obstacles to lung cancer screening include a lack of awareness, cost and coverage concerns, and a lack of physical access (15).

There are many risk factors associated with a lung cancer diagnosis. Some of these include firsthand smoke exposure, secondhand smoke exposure, exposure to toxic gases such as radon, and exposure to environmental pollution such as asbestos (16). With two major factors being air pollution and secondhand smoke, a large research focus is on understanding how pollution is spread geographically and how pollution levels influence cancer incidence (17;18;19). There is clear evidence that polycyclic aromatic hydrocarbons (PAHs) which are primarily emitted from traffic exhausts and fuel combustion increase the risk of lung cancer (20). Also, air pollution in the form of PAHs has been shown to be positively geographically correlated with coal-fired power plants (21). Therefore, the workers and individuals who are located near power plants are being exposed to PAH pollutants at a higher rate than others. A study has shown that coal-fired power plant workers have an increased cancer risk due to airborne PAH inhalation and dermal contact (22). Also, it is well known that people living near powerplants suffer from higher rates of premature mortality, respiratory
diseases, lung cancer, cardiovascular disease, poorer child health, and higher infant mortality (23). Within this conceptual framework, our research team believed it would be critical to understand how the coal-powered plants impact the lung cancer incident rates across Kansas (24). The state of Kansas contains 105 counties which are spread across 82,278 mi² and a state population of 2.93 million. The state of Kansas contains less air pollution relative to other highly populated states in the Midwestern United States (25).

Coal-powered plants are in many different counties across the state of Kansas. To relate those power plants with county lung cancer rates requires the usage of a geo statistical tool such as ArcGIS (26). Our primary research aim was to conduct a spatial analysis evaluating whether counties that have a coal-powered plant had a higher lung cancer incidence compared to counties that do not have a coal-powered plant.

**Method:**

Throughout Kansas, there are coal-powered power plants. The impact of the coal-powered power plants was evaluated using a spatial join to check the relationships between lung cancer incidence against lung cancer rates, comparing to see if Kansas counties within the smoke range of coal-powered power plants differ from those of Kansas counties outside the smoke range. The spatial join operation is the affixing of data from one feature layer attribute table to another from a spatial perspective; the process is called a spatial join which can be performed using the Geographic Information System (GIS). First, we built a layer in ArcGIS with just the lung cancer incident rates. Then we used the National power plant data set to identify what coal power plants were in Kansas which was added as a new layer. To join the first and second layers we used a spatial join. The spatial join outcome was used to examine areas that would be influenced by the dissipation of smoke emitted from these power plants. For a secondary analysis, we created a generalized linear model (GLM) to understand better the coal-powered power plants’ impact on surrounding areas. The GLM captured the relationships between lung cancer rates and how the rates are affected by lung cancer risk factors, including the dissipation of smoke from the plants in the surrounding counties.

Data Source: For this analysis, we utilized two distinct datasets that were associated using spatial joining. The first dataset was the lung cancer incident rate data set for five years (2010–2014). The source of this data set is ArcGIS repository The unit of measure within this data set is county level (105 counties across the state of Kansas). The lung cancer incident rates within the data set are age-adjusted. An age-adjusted rate is defined as the rate that would have existed if the population under the study had the same age distribution as the standard population (27). Additionally, the data set included smoking rates, total population, percent of the population below the poverty level, and percentage of people between the age of 55 to 79 for each county. The incident rate data was mapped in ArcGIS, as shown in Fig. 1.

**Figure 1:** Lung cancer incident rate per 100,000 people for the state of Kansas, five-year period 2010–2014

**GLM Model:**

\[ Y_i = -10686.634 + 247.664X_1 + 0.05X_2 + 315.048X_3 - 0.003X_4 \]

The second data set that was utilized was sourced from the ArcGIS online – which is "minn_2020_Power_Plants". This data set included information such as Primary Fuel source, location of the power plant, plant name, utility name, geographical location, and the amount of energy generated from these power plants. The primary Fuel Source variable details if the power plant uses coal, hydroelectric, wind, petroleum, biomass, solar, batteries, geothermal, or other fuel sources to generate electricity. Based on our research goals for this study we restricted the power plant data set to only include coal plants located within the state of Kansas. Five coal power plants were identified across the state of Kansas as shown in Fig. 2.
Figure 2: Map highlighting the coal-powered plants that are geographically located across the state of Kansas.

The details of the coal-powered plants that are located within the state of Kansas are detailed in Table 1 as shown below:

<table>
<thead>
<tr>
<th>Plant Name</th>
<th>Utility Name</th>
<th>County</th>
<th>City</th>
<th>Zip</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jeffrey Energy Center</td>
<td>Evergy Kansas Central, Inc</td>
<td>Pottawatomie</td>
<td>St. Mary's</td>
<td>66536</td>
</tr>
<tr>
<td>Lawrence Energy Center</td>
<td>Evergy Kansas Central, Inc</td>
<td>Douglas</td>
<td>Lawrence</td>
<td>66049</td>
</tr>
<tr>
<td>Nearman Creek</td>
<td>City of Kansas City - (KS)</td>
<td>Wyandotte</td>
<td>Kansas</td>
<td>66104</td>
</tr>
<tr>
<td>La Cygne</td>
<td>Evergy Metro</td>
<td>Linn</td>
<td>La Cygne</td>
<td>66040</td>
</tr>
<tr>
<td>Holcomb</td>
<td>Sunflower Electric Power Corp</td>
<td>Finney</td>
<td>Holcomb</td>
<td>67851</td>
</tr>
</tbody>
</table>

Table 1: Coal Powered Plant located across the state of Kansas

Analysis: The analysis was comprised of a few steps. The first step was determining how many coal power plants are within the Kansas borders. Once the information was found (Table 1) there needed to be a secondary search conducted to determine which county zip codes were within a predefined radius of the five power plants; but first we had to determine what that predefined radius of the study would be.

The radius was determined to be 25 miles based on a few critical factors in the research done by Morehouse & Rubin, Yang et al., and Jha & Muller. We considered the air pollution, the effect of pollution on pregnant mothers, and emissions from coal stockpiles at power plants. First, air pollution can travel long distances, often reaching beyond state borders (28). Second, research has shown pregnant women living 20 to 30 miles downwind from a power plant are at greater risk of having a child who suffers from a low birth rate (29). Finally, coal stockpiles at powerplants have been shown to increase the emission of fine particulates by 10%, causing a 0.09% increase in fine particulates within a 25-mile radius of the power plants (30). Based on the three factors above, we decided on setting the study radius around the Kansas power plants at 25 miles. Using a 25-mile radius from each power plant, we could determine which Kansas counties fell within the preset range. The next step was to test the association between the distance from the powerplant predictor and the outcome variable lung cancer incidence using spatial join.

We tested the association between the predictor of distance from a power plant against lung cancer incidence using the spatial join mentioned earlier. The spatial join was conducted using our power plant buffer layer and the lung cancer incidence layer to establish the spatial relationship between these two layers of interest. The analysis of the control variables’ effects on lung cancer incidence was also conducted as a secondary analysis by using a GLM to explore their significance and predictive capability. We analyzed the various control variables of percent current smokers, percent population between 55 and 79, and percent below poverty by county.

Included beneath Fig. 1 is the GLM that was created to perform the secondary analysis of the study. The GLM was used to analyze the various relationships between the outcome variable, lung cancer incidence, and the control variables: distance to coal plant, percent current smokers, percent population between 55 and 79, and percent below poverty. The GLM was used to give vital data included in Table 3. With that information, we could determine which control variables had the most significant impact on lung cancer incidence and make vital conclusions based on the GLM.
Table 3
Summary of GLM using Gaussian Model.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>t-Statistic</th>
<th>Probability</th>
<th>Robust Standard Error</th>
<th>Robust t-Statistic</th>
<th>Robust Probability</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-10686.634</td>
<td>3933.389</td>
<td>-2.717</td>
<td>0.008</td>
<td>3180.645</td>
<td>-3.360</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Percent Current Smoker (P_CUSMK)</td>
<td>247.664</td>
<td>299.804</td>
<td>0.826</td>
<td>0.411</td>
<td>242.679</td>
<td>1.021</td>
<td>0.301</td>
<td>2.063</td>
</tr>
<tr>
<td>Percent Population between 55 and 79 (POP_55TO79)</td>
<td>0.050</td>
<td>0.023</td>
<td>2.206</td>
<td>0.030</td>
<td>0.0139</td>
<td>3.594</td>
<td>0.001</td>
<td>1.013</td>
</tr>
<tr>
<td>Percent below Poverty (P_B_POV)</td>
<td>315.048</td>
<td>129.541</td>
<td>2.432</td>
<td>0.017</td>
<td>110.835</td>
<td>2.842</td>
<td>0.005</td>
<td>2.056</td>
</tr>
<tr>
<td>Nearest Distance (NEAR_DIST2)</td>
<td>-0.003</td>
<td>0.005</td>
<td>-0.844</td>
<td>0.401</td>
<td>0.004</td>
<td>-0.866</td>
<td>0.388</td>
<td>1.014</td>
</tr>
</tbody>
</table>

Results:

The geospatial join between the lung cancer incident rate and the coal-fired power plant location demonstrates that the counties with no coal-fired power plant and the county that has a single coal-fired power plant weren't too different regarding the lung cancer incident rates. The counties that fell under the area of influence of the two coal power plants had a slightly higher lung cancer incident rate as shown in Fig. 3 below compared to counties with zero.

Figure 3: Boxplot representing the relationship between Coal Power Plants and Lung Cancer Incident rate per 100k

The median for the counties with zero coal plant is 53.55 and the Interquartile Range (IQR) is 19.58; counties with one coal plant had a median of 59.1 and an IQR of 12.8 the counties that fell under the influence of two coal plants had a median of 68.45 and an IQR of 8.67. The small IQR and higher median suggest that the lung cancer incident rates for the counties that fall under the area of influence of two coal plants may be higher than counties with zero. Counties that are categorized under zero area of influence have outliers and a larger IQR suggests that the incident rates are not consistent across these counties.

The correlation between the variables used in the generalized linear model analysis is shown below in Fig. 4. Smoking rate, population above 55 to 79, population below poverty, and the number of oncology hospitals had a positive relationship with the lung cancer incident rate, the nearest distance had a negative relationship with the lung cancer incident rate. This is what we had expected i.e., the counties that are farthest away from one of these coal-powered plants will have a smaller lung cancer incident rate.

Figure 4: Correlation between variables – GLM analysis, Model - Gaussian

Summary statistics for the variables are presented below in Table 2 and the generalized linear model analysis summary results are presented below in Table 3.
Table 2: Summary of GLM using Gaussian Model.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>Max</th>
<th>Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung Cancer Incidence (L_Id_14)</td>
<td>55.44</td>
<td>15.44</td>
<td>97</td>
<td>15.2</td>
</tr>
<tr>
<td>Percent Current Smoker (P_CUSMK)</td>
<td>16.11</td>
<td>1.81</td>
<td>22.96</td>
<td>11.99</td>
</tr>
<tr>
<td>Percent Population between 55 and 79 (POP_55TO79)</td>
<td>27.06</td>
<td>4.86</td>
<td>37.27</td>
<td>11.40</td>
</tr>
<tr>
<td>Percent below Poverty (P_B_POV)</td>
<td>12.53</td>
<td>4.17</td>
<td>23.2</td>
<td>4.3</td>
</tr>
<tr>
<td>Nearest Distance (NEAR_DIST2)</td>
<td>77.36</td>
<td>51.32</td>
<td>187.60</td>
<td>0</td>
</tr>
</tbody>
</table>

The population below poverty and the population between the ages of 55 and 79 years are the two factors that have statistical significance in the model, the other factors such as smoking rate, and nearest distance do not depict statistical significance, based on their P-values at an alpha level 0.05.

Figure 5: Lung Cancer Incidence Standardized Residuals

Looking at incidence predictions by county, we notice that southwestern Kansas has a cluster of counties with negative standardized residuals while northwestern Kansas has positive standardized residuals. Notable about southwestern Kansas is the presence of only one coal power plant, however, it should be noted that the sparse and small population of western Kansas may be impacting the model predictions.

Discussion:

There is a wealth of literature evaluating smoking as a risk factor for cancer, however air pollution has comparatively little research as a risk factor. Given the concentration of air pollutants in urban areas with high population concentrations, it should be a far greater concern for overall health.

Coal-fired power plants emit roughly 2,180 pounds of carbon dioxide per megawatts of electricity produced. During the summer months of 2019, the state of Kansas generated 50,888 gigawatt-hours of electricity, of which 34% was generated by coal-fired plants (6). While the overall reduction of operational coal-fired plants in the past decades has been an indication of progress, it is important to not allow that progress to stagnate. However, the question remains if alternative power sources will be sufficient to provide the remaining 34% of power generation required by the state.

Additionally, can we remediate excessive toxic emissions by growing and maintaining ecosystems around the power plants? Trees act as carbon sinks, but if they also store the other harmful emissions then the negative impacts of these powerplants can be reduced by expanding surrounding forests. Also, cities should ensure that the land around the coal-fired power plants is not sanctioned for developing housing communities and hospitals as it may pose an immediate hazard. Government policy should also be examined and written to make sure it is providing a safe work environment for workers at coal power plants.

Limitation:

Due to the lag of cancer reporting methodology, the lung cancer rates are not recent. These rates could be slightly different based on the recent lung cancer detection. We still believe the demographics and the coal plants spread across
the state of Kansas have remained the same so this should not impact our overall findings. Neighboring states might also have a coal plant located close to the Kansas border which might be contributing to the air contamination. Polycyclic aromatic hydrocarbon levels are not recorded in the dataset we have and therefore could not specifically be investigated in this study. Other climatological factors such as wind, rain, and humidity were not accounted for in our study, as a future study we plan to gather this data and ensure it is included in the regression model. There could also be a temporal lag due to the nature of the disease as cancer often takes years to develop. People may also move housing locations during this period which could also change their exposure to pollution. Demographic background smoking rates are not accounted for in this study. Assuming low-income individuals live closer to the power plants the increased smoking rates among that population certainly have an impact on lung cancer incidence as well. Low-income smokers also tend to smoke more than wealthier smokers and are more dependent on nicotine (10). Because of how correlated smoking, lung cancer, and poverty are this is a limitation of the study.

Conclusion:

Based on the lung cancer incident rate distribution across the state of Kansas, it seems like the metropolitan cities experience a higher rate compared to the rural areas. The age-adjusted lung cancer incident rate in Kansas (55 per 100k) is however higher than that of the National average (52 per 100k). There wasn't any statistical significance on the distance from the coal-fired power plant, but further investigation is necessary.

Based on the box plots, the counties near one coal-fired power plant do not have a higher lung cancer incident rate compared to the counties that have no coal power plant, but the counties that are close to two plants may have higher rates compared to counties near no coal power plant.

The regression analysis has demonstrated that a relationship exists among some of the predictors used with Percent Population between 55 and 79 and Percent below Poverty being significant predictors. Further analysis and a more comprehensive dataset must be used to identify in detail how lung cancer incident rates may be impacted due to coal-fired power plants.

Our analysis has indicated that lung cancer rates may be higher in the areas around two coal-fired power plants, however, our analysis lacks other factors which could be important for a comprehensive analysis. Future analysis should be done to examine how Kansas’s coal power plants are impacting the workers who work there as they are likely exposed to the highest level of pollutants. How do their lung cancer incidence rates compare to that of the general population?

Declarations

Ethics approval and consent to participate:

The IRB at the University of Kansas Medical Center has waived the need for approval for this study as this is considered as a quality improvement study.

Consent for publication:

Not applicable

Declarations
Competing interests:

The authors declare that they have no competing interests

Funding:

This study was supported by the National Cancer Institute (NCI) Cancer Center Support Grant Funding P30 CA168524 and used the Biostatistics and Informatics Shared Resource (BISR).

Authors' contributions:

DM analyzed and interpreted the data. DM, SP and AA assisted with background research, figure and model creation, citations, and writing. BG assisted with statistical methodology and interpretation. All the authors were involved with drafting the manuscript.

Acknowledgements:

Not applicable

References


7. 7.</number>


10. 10.</number>


**Unsectioned Figure Details**

A
Fig. 4
Correlation between variables – GLM analysis, Model - Gaussian

A
Fig. 5
Lung Cancer Incidence Standardized Residuals

A
Fig. 1
Lung cancer incident rate per 100,000 people for the state of Kansas, five-year period 2010–2014

A
A
Fig. 2
Map highlighting the coal-powered plants that are geographically located across the state of Kansas.

A
Fig. 3
Boxplot representing the relationship between Coal Power Plants and Lung Cancer Incident rate per 100k

Unsectioned Paragraphs

Word Count: 2147

Abbreviations:

- Polycyclic Aromatic Hydrocarbons (PAHs)
- Geographic Information System (GIS)
- Generalized Linear Model (GLM)
- Interquartile Range (IQR)

Figure Legends:

Figure 1: Lung cancer incident rate per 100,000 people for the state of Kansas, five-year period 2010–2014

Figure 2: Map highlighting the coal-powered plants that are geographically located across the state of Kansas.

Figure 3: Boxplot representing the relationship between Coal Power Plants and Lung Cancer Incident rate per 100k

Figure 4: Correlation between variables – GLM analysis, Model - Gaussian

Figure 5: Lung Cancer Incidence Standardized Residuals