A Spatial Analysis of Mining Activity and Cancer in Montana Counties

Harvard College, Class of 2007
ES 103 Final Project

Introduction

The mining industry is responsible for several detrimental effects on health and the environment through the mining process itself as well as through pollution.\(^1\) Although there is a wide range of health impacts associated with mining depending on the mineral and mining type, dust is common to all types of mine processes, and respiratory effects, especially for workers, are among the most serious and most widely researched.\(^2\) Most studies on mining and health have focused on mineworkers with relatively fewer studies on health effects on general members of the population living in the vicinity of mines because of the difficulty in undertaking adequate studies to assess possible relationships and the uncertainty and controversy that still remains among different interest groups over the actual existence of community effects on health.\(^3\)

Spatial analysis would provide a simpler way of approaching a community-level assessment of mining’s health impacts in the absence of studies on individuals. Furthermore, since geologic features—such as soil type and mineral deposits—with similar characteristics are not randomly distributed in space but tend to be situated together, there may also be a spatial component to potential health effects resulting from mining. Because of the prevalence of respiratory problems involved in the mining industry, this study focuses specifically on lung cancer as a potential health problem related to mining in order to determine if there are spatial differences in disease rates in areas with different levels of mining activity. Additionally, cancer is a significant illness because it can surface much later after exposure, and there is still a lack of sufficient information on long-term as opposed to more direct respiratory illnesses related to mining.\(^4\) In this study, the spatial examination of the relationship between lung cancer and the location of mines will be applied to the state of Montana because of its history as a mining state and the readily available spatial data for its mining operations.

Data

Spatial data was obtained from the Montana Geographic Information Clearinghouse on their Natural Resource Information System website (http://nris.mt.gov/gis/default.asp). Shapefiles used for this study included a polygon layer of counties as well as population data from the file 2000 Census County Divisions. Mining data came from the point shapefile Abandoned and Inactive Mines Database, which is claimed to be the most comprehensive listing

---

4. Ibid, 27.
of Montana mines.\(^5\) A mine is classified as abandoned if its owner cannot be identified or if it is federally owned, and an inactive mine has an owner but is not presently running.\(^6\)

Cancer data was downloaded from the National Cancer Institute State Cancer Profiles website (http://statecancerprofiles.cancer.gov/) as Microsoft Excel files. In particular, a data table for lung and bronchus cancer incidence rates by county for all races, both genders, and all ages was used for the analysis. This data is age-adjusted and provides annual incidence rates per 100,000 people for the period 2000-2002, which is the most recent available; however, for counties with fewer than 16 cases of cancer, rates are not provided but suppressed for confidentiality reasons and because rates based on small numbers are not statistically sound.\(^7\) A similar data table was obtained for incidence rates of all cancer in Montana counties. In addition, demographic data based on the 2000 census that seemed to be likely to influence exposure to mining risks was obtained from this website. This included a table of the percentage of the population in each county that has not moved in the last five years because longer residence would result in higher exposure to health risks; and—although there was no county data available for percent of the population employed in the mining industry to factor in to the analysis—a table with the percent of the county populations occupied in white collar jobs was used since the indoor nature of the work may limit exposure to mining health risks.

The lung and bronchus cancer incidence rates, occupation, and mobility data were added to the county polygon shapefile attribute table in order to reference this data spatially and produce quantile maps, and a spatial join of the mine layer on the counties was conducted to produce a count of mines in each county, which was then converted to a density for easier comparison by dividing by the county area in square kilometers. This process was repeated for new layers created with mines selected based on specific attributes considered to potentially be related to lung cancer effects. These included mines whose status is still listed as producing and mines with surface operations because, in comparison to underground mines, studies on surface mines center on dust and respiratory problems.\(^8\) Mines extracting the commodities coal and copper were also selected because of the known dangers of coal dust, and most research on coal mining and health now focuses on dust and respiratory illnesses; bronchitis, silicosis, pneumoconiosis, and lung failure have all been attributed to working in coal mines.\(^9\) Furthermore, many studies have found increased mortality from lung cancer and other respiratory illnesses in communities near copper mines.\(^10\) The population data from the 2000 Census County Divisions shapefile was dissolved to the county level and joined to the county and mine data to calculate density of mines per person in each county as another measure for comparing mining activity.


\(^8\) Stephens and Ahern, 24.

\(^9\) Ibid, 18.

\(^10\) Ibid, 38.
Exploratory Spatial Data Analysis (ESDA)

Mapping the mines on the county cancer rates does not show clear relationships for either lung cancer or all cancers based on visual observations, and counties with the some of highest mine density with area—those in the southwest corner of the state—do not necessarily have high lung cancer rates, as the white color indicates for counties with suppressed lung cancer rates (Figures 1 and 2). The counties with the highest lung cancer rates, those in black, appear to be dispersed. Similarly, no clear link is seen from maps of only surface mines or producing mines, but a map of the commodities—coal and copper—seems to be indicative of some relationship to county lung cancer rates (Figure 3). The clustering of copper mines along the western side of the state and the two clusters of coal mines in the south and the northeast do not appear to necessarily be directly in the counties in the highest quantile of lung cancer incidence rates but surrounding them, and perhaps factors such as the prevailing wind direction could influence the expression of lung cancer in the nearby counties.

To further investigate possible relationships, GeoDA was used to produce scatterplots and test for spatial autocorrelation. No clear trends are evident in the scatterplots of mine count and density versus lung cancer or all cancer rates for all mines—as seen in the clustered points in the scatterplots (Figure 4)—or for surface mines, producing mines, copper mines, and copper and coal mines combined. Plotting the percent white collar and percent of people living in the county for more than five years also does not display a clear trend. However, the strongest correlation is visible in the scatterplot with coal mine density and lung cancer rates (Figure 5).

The trend lines drawn in the plots are also influenced by the counties with suppressed cancer rates, which are all plotted as zero although the rates are not necessarily zero. After excluding these 31 of the 57 counties that had suppressed lung cancer rates—although it alters the trend lines, even changing directions in the plot of all mine density and lung cancer—since the remaining points are still grouped in the same way, the lack of clear correlations persists. Although none of the trends are strong based on the clustered points, for mobility and occupation, the trend becomes positive with lung cancer excluding these counties, which is expected for mobility but indicates an unexpected increase of lung cancer with more office
workers. For the density of all mines and producing mines, the plots have unexpected negative relationships with lung cancer rates but seem to be influenced by outliers. For surface and coal mines excluding suppressed counties, the trend lines become effectively horizontal, indicating no relationship.

Figure 3: Mines with coal or copper production plotted with lung cancer incidence rates

Figure 4: Scatterplots of total mine count and density and mobility and occupation versus all cancer rates (top row) and lung cancer rates (bottom row)
Spatial autocorrelation was examined using the Moran’s I and local indicator of spatial association (LISA) statistics using a weights file based on the four nearest neighbors as well as a queen’s contiguity weights file. As expected, there is significant positive spatial autocorrelation ($I=0.7493$) for total mine density in counties, with counties with high densities of mines in red clustered in the southwest and counties with low mine densities in blue clustered in the east (Figure 6). Similar spatial autocorrelation is found with the surface, producing, and copper mines. The location of coal mines also has a significant but weaker positive spatial autocorrelation ($I=0.1269$), and their distribution differs, with low density counties in the southwest, and a high density county in the south and the northeast. In general Moran’s I values for the mine variables with the cancer rates are insignificant and low or near zero both with and without the counties with suppressed cancer rates, but the queen’s contiguity matrix produces slightly stronger autocorrelation results. There is effectively no significant spatial autocorrelation of counties based on cancer incidence rates although it is slightly negative for lung cancer ($I=-0.0399$) and slightly positive for all cancers (0.0319). With the queen’s contiguity, on the other hand, lung cancer has a slight positive spatial autocorrelation. There is a slight negative but insignificant spatial autocorrelation for all mines and lung cancer with a Moran’s $I = -0.26$ excluding the suppressed counties. Counties with high lung cancer rates are surrounded by counties with both high and low mine densities. For coal mines excluding suppressed counties, there is a positive yet insignificant spatial autocorrelation with lung cancer rates ($I=.1686$) again with high lung cancer counties surrounded by counties with both high and low coal mine densities. Surface mines have a negative autocorrelation with lung cancer ($I=-0.2414$) which is not statistically significant, and counties with high lung cancer are mainly surrounded by counties with low density of surface mines as well as counties with high surface mine densities.
This is the same relationship with the producing mines and lung cancer, with $I=-0.2529$. For density based on mines per person, the results are similar in that there are not clear trends in the scatterplots but slight positive trends in surface, copper, and particularly coal mines, and a slight negative trend in producing mines. The spatial autocorrelations excluding the counties with suppressed cancer rates are comparable and perhaps stronger, with the strongest relationship being a positive autocorrelation between copper mine densities and lung cancer rates in surrounding counties ($I=0.4268$), although, in its Moran’s I plot, the data observations are clustered around the origin (no spatial autocorrelation) with just a few outliers in the high copper mine density surrounded by high lung cancer incidence quadrant.

**Principle Components Analysis**

Rather than running a regression to model how the selected variables are related, because of the unclear ESDA results, Principle Components Analysis (PCA) was instead performed using SPSS. PCA only determines which variables are most related by producing independent vector components that explain the greatest amount of deviation in the original variables in order to simplify the number of variables. An initial PCA on all of the variables selected from the data—lung cancer rates, all cancer rates, mobility, occupation, total mine density, surface mine density, producing status mine density, copper mine density, and coal mine density—extracted three principle components, the first being correlated with high mine density for all mine variables except coal mines, similar to the ESDA results in which all the mine variables displayed a similar spatial pattern except the coal mines. The second component is highly correlated with demographic variables—mobility and occupation—meaning that it is explained by simultaneously high levels of office workers and high levels of long residence in counties; and the third component is correlated with high cancer rates and, in addition, somewhat high coal mine density. This means that there is a possible relationship between coal mine density and cancer rates or that these variables behave similarly with this component.
However, because of the relatively small amount of data, 57 counties, PCA with this number of variables may not be reliable, and it is suggested that there should be no more than one variable per ten observations. Furthermore, because the mine density variables could be related to each other, including fewer of them in the analysis could improve results. Therefore, additional PCAs were conducted using different combinations of variables, and the best result—for the five variables lung cancer rates, coal mines, surface mines, all mines, and mobility—extracted three components that explained 86.083% of the total variance and a high proportion of the variables’ variances, as indicated by the high values in the extraction column of the communalities table (Table 1). The component matrix (Table 2) indicates the correlation between the variables and the components, where large positive or negative values close to 1 indicate strong correlation. Again, the first component is described by mine density, as the large values in the matrix indicate, and the second has a strong correlation with coal mine density and lung cancer incidence. The last component is mainly explained by the mobility variable, but the coal mine variable is also moderately correlated with this component. Mapping the scores for the components shows a cluster of high component 1 scores in the southwest where the counties with high mine densities are located (Figure 7). The coal mines are mapped with the second component scores (Figure 8) to show the correspondence, and the counties with higher scores correspond to the counties with higher lung cancer incidence mapped above. The third component scores are also mapped with the coal mines since it also is rather high on this component. A map of mobility is depicted beside the score map for comparison (Figures 9 and 10).

The fact that the coal mine variable is also relatively high on the third component is not desirable in PCA analysis since the components are unrelated and variables should be highly correlated with just one factor. However, other PCA results for different variable combinations either also have a variable with somewhat high correlation in more than one component, or the extracted components do not explain the variance in the data as well. Regardless, the high correlation with lung cancer and coal mine density is a consistent result in the extracted components, as well as the results found in the initial PCA with mine densities correlated, the mobility and occupation variables correlated, and the cancer rates correlated. Other interesting common results after running additional PCAs are that some components have mobility correlated positively with lung cancer, as anticipated, but other PCA results indicate the opposite relationship. There are also a few PCA results with coal mine density and percent white collar workers correlated highly on a component, which is not anticipated since one would expect counties with more coal mines to have more mine workers and thus fewer office workers. Therefore, this result would mean that high lung cancer incidence is not just attributed to miners themselves but to other members of the community. Furthermore, there are PCA components with lung cancer rates and surface mines highly correlated inversely, implying unexpectedly that counties with more surface mines have lower lung cancer rates.

The PCA results with densities based on population rather than area indicate mostly similar relationships with the coal mines and lung cancer rates as well as other factors. When excluding the suppressed data, the PCA results are generally poor, possibly because of the use of only a few variables since there are only 26 data observations, which often resulted in only one or two extracted components, although including more variables led to better results. However,

12 Ibid.
these PCA results do not express the same strong correlation with coal mine density and the lung cancer variable, although the two variables are still both correlated positively.

<table>
<thead>
<tr>
<th></th>
<th>Initial</th>
<th>Extraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>LungAnnInc</td>
<td>1.000</td>
<td>.707</td>
</tr>
<tr>
<td>minedens_coal</td>
<td>1.000</td>
<td>.797</td>
</tr>
<tr>
<td>minedens_surf</td>
<td>1.000</td>
<td>.964</td>
</tr>
<tr>
<td>MINEDENS</td>
<td>1.000</td>
<td>.964</td>
</tr>
<tr>
<td>MOBILITY</td>
<td>1.000</td>
<td>.872</td>
</tr>
</tbody>
</table>

Extraction Method: Principal Component Analysis.

<table>
<thead>
<tr>
<th></th>
<th>Component</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>LungAnnInc</td>
<td>.193</td>
</tr>
<tr>
<td>minedens_coal</td>
<td>-.083</td>
</tr>
<tr>
<td>minedens_surf</td>
<td>.958</td>
</tr>
<tr>
<td>MINEDENS</td>
<td>.962</td>
</tr>
<tr>
<td>MOBILITY</td>
<td>-.286</td>
</tr>
</tbody>
</table>

Extraction Method: Principal Component Analysis
a 3 components extracted

Figure 7: Principle component 1 scores

Figure 8: Principle component 2 scores
Discussion

The results of the PCA indicate that there is likely a positive link between coal mining and lung cancer incidence, which is supported by previous findings; however most of these findings have focused on mineworker health while this indicates a similar link in communities in general near coal mining activity. However, since the suppressed data does not necessarily imply that there is no incidence of lung cancer in those counties as is assumed with their inclusion, analysis excluding these counties should be more accurate. Yet, the relationship did not hold when the data was analyzed without the suppressed rates, which may indicate that these counties were functioning as outliers. It may also be the case that the suppressed rates actually are the lowest lung cancer incidence rates since they are counties with below 16 lung cancer cases, and if the populations of these counties themselves are not also very small, then the relationship expressed including these counties may be rather true. The results could perhaps be explained by the possibility that—as with many typical dose-response curves for health hazards—there may be a threshold for the amount of coal mining activity needed to cause a sizeable response in lung cancer rates to move counties out of the suppressed range rather than a continuous relationship starting from any mine density above zero. Furthermore, there could also be a saturation point or upper limit after which increasing mine density will not make much of a difference in lung cancer incidence. The lack of correlation without suppressed counties might be due to saturation which is more easily detected when this data is excluded. In addition, the small sample size of 26 for the remaining counties means that the results could be highly influenced by outliers or high variance in the data, which could have also prevented a relationship from being ascertained.

An interesting unanticipated result is that there are no strong relationships between other mine factors such as active mines, copper mines, or surface mines with lung cancer incidence. Possible reasons could be the way in which the data was analyzed. For instance, in the selection of mining commodities, it may have been more appropriate to examine only the active coal or active copper mines as well as just the active surface mines. Although the mines with active status display a similar spatial pattern as most of the other variables on mine density, the south-
westernmost county, Beaverhead County, is significantly negatively spatially autocorrelated with other counties for this variable. It has a low density of producing mines although it is surrounded by counties with high densities of active mines—as the light blue color indicates in the cluster map—and although it also has a relatively high density of mines overall (Figure 11). The observed results could also be influenced by the completeness of the data itself. The metadata for the mine shapefile indicates that it is based mainly on the US Bureau of Mines Minerals Industry Location System (MILS) Database in addition to other sources reporting mine locations. However, only mines contained in the MILS database have data on the commodity, status, and type of operation, all of which were variables in this analysis. 1272 of the 8449 mine records do not have information on these attributes. In terms of the cancer data, incomplete recording and differing levels of obtaining cancer information in counties could cause the observed results to differ from actuality. Places that have populations who are more involved in health care services, such as undergoing testing for cancer, would by default report higher incidence rates; and chance could come into play in terms of when the disease is finally expressed or detected. 13

Despite these possible complications, the results of this analysis could still be reliable. Although there are conclusive findings on mineworker health in terms of lung problems, some community-based studies also did not find significant higher rates of actual respiratory disorders in people living near surface coal mines, although there was higher reporting among children of respiratory symptoms. 14 Similarly, there have been mixed results dealing with whether surface mines are associated with higher respiratory problems in mine workers. One 1977 study did not find much support for such a link, while some later studies did find that surface mine jobs involving a lot of dust led to higher respiratory and oral complications. 15 In terms of copper mining, although the previous studies mentioned earlier in the data section found significant higher levels of lung cancer, these were based on mortality rates rather than incidence. Furthermore, one of these studies, a 1975 US study, found significant higher lung cancer mortality rates for Caucasians, and a 1983 Quebec study found significant results in men for increased lung cancer death and for other illnesses in women. 16 Thus, although age has already been adjusted for in the cancer data used in this study, further research could factor in other demographic variables such as race and gender as well as analyzing mortality; however, the data available for cancers selecting for these specific factors is less complete and contains more suppressed counties.

Temporal analysis is likely to also be significant for studies on cancer and mine activity. Examining cancer rates over earlier time periods or cumulative rates could be a good way to see how the relationship has changed with time and if there are stronger links with mining activity for earlier time periods. Similarly, the years that mines have been in or out of operation could also influence cancer rates since cancers can develop after many years and since the strength of residual effects remaining in the environment could decline with increased time from the ceasing of mine operation. In addition, other minerals could also be analyzed, such as asbestos and uranium mining, which are also linked to many respiratory problems. 17

Finally, in terms of lung cancer, there are other significant causes that may exhibit different levels in different counties. For instance, smoking is the number one cause of lung

---

14 Stephens and Ahern, 30.
15 Ibid, 25.
16 Ibid, 38.
17 Ibid, 8.
cancer. Thus, this behavioral factor would be important to control for in analyses since smoking rates may differ with location. However, smoking data for Montana is only collected at the state level and not the county level. Secondly, indoor air quality is also important since, after smoking, radon—a radioactive gas which is found both indoors and outdoors and results from uranium in the environment—is the next leading cause of lung cancer. The EPA has a map of radon zones in counties (Figure 12), but it is not very detailed but based only on a three level system. Most of Montana has the highest potential for high indoor radon concentrations (more than 4 pico curies per liter), except for some counties with moderate potential (2-4 pCi/L), most of which also have suppressed lung cancer incidence rates. Exploring other factors related to the physical environment that may influence the distribution of risks from mining activity—such as winds, the slope of land surface, and the flow of rivers—may be also useful.

Figure 11: LISA cluster map for density of producing mines
Figure 12: EPA Map of Radon Zones
Source: [http://www.epa.gov/radon/zonemap.html](http://www.epa.gov/radon/zonemap.html)

Implications

If coal mining is indeed affecting community health as some of the results of this study seem to indicate, this has policy implications for affected areas. It may suggest the need to create buffers of a certain distance or area around past and present mines to limit human activity. For example, another less comprehensive mine shapefile from the Montana Department of Environmental Quality contained an attribute for land use for mines, which included urban, residential, recreation, built, and agricultural land, suggesting that some mines—even if no longer active—are located in zones with human activity. Environmental justice may also come into play. Although no other mining characteristics were strongly associated with lung cancer incidence, the fact that there is spatial autocorrelation for mine concentration and that, for instance, the majority of active mines are clustered in the southwest raises questions about what the population in this area is, in terms of its size, ethnicity, and income level; and this would need to be investigated to suggest measures to alleviate any potential excess risk placed on those living in the vicinity of these mines. It would also be useful to compare the cancer rates in specific Montana counties with average US or statewide rates to determine which counties exceed this average and may require more priority attention in terms of policy efforts in public health.

---

19 Ibid.