A GEOGRAPHICAL AND STATISTICAL ANALYSIS OF LEUKEMIA DEATHS RELATING TO NUCLEAR POWER PLANTS

Whitney Thompson, Sarah McGinnis, Darius McDaniel,

Jean Sexton, Rebecca Pettit, Sarah Anderson, Monica Jackson

ABSTRACT: Due to an alarming increase in childhood leukemia rates, it has become a priority to learn what factors contribute to this disease. While no direct cause is known, it has been suggested that environmental factors such as the radiation emitted by nuclear power plants may be to blame. This study examines whether or not there is a geographical pattern in cancer rates based upon the locations of nuclear power plants throughout the United States. A spatial analysis was conducted to look for global and local clusters with an increased mortality rate due to childhood leukemia along with a statistical analysis to examine which factors contribute significantly to this rate. We found no evidence to prove that nuclear power plants are responsible for the recent increase in childhood leukemia rates. This leads us to believe that a different carcinogen is at fault.

Key words: statistics, childhood cancer, linear regression, Tango's Index, Moran's *I*, Oden's *I*pop*

I) INTRODUCTION

Childhood leukemia rates have risen 39.6 percent between 1975 and 2000 [9]. While direct causes of leukemia are unknown, it is believed that environmental factors, such as radiation emitted from nuclear power plants, may play a role in the growing number of new cases each year. This alarming increase has led to a number of studies examining the correlation between these leukemia rates and the distance the sick child lived from a nuclear power plant.

Leukemia, a type of cancer found in bone marrow, is one of the types of cancer most commonly affected by radiation [9]. Leukemia occurs when a cell begins to divide irregularly. This irregular cell division can lead to the formation of uncharacteristic chromosomes. Exposure to ionizing radiation increases the likelihood that transformations of the chromosomes occur [8]. The isotopes known as strontium-89, strontium-90, and barium-140 commonly target bones and bone-marrow. These cancer causing isotopes, emitted by nuclear power plants, are not part of naturally occurring forms of radiation.

For this reason, we will be studying whether or not leukemia rates in counties containing a nuclear power plant differ from those without one.

Safety has always been an issue concerning nuclear power plants. There are many Americans who feel that nuclear power plants are not safe, forcing many plants to be shut down over the years. These safety concerns have also led the government to become more involved in the process of building and updating the plants and to become very strict about enforcing regulation rules for nuclear power plants. The Nuclear Regulatory Commission (NRC) is in charge of overseeing that all plants are working properly and are up to date. The safety regulations and rules involve operation, accident prevention, and emergency plans [3]. NRC regulation helps to make sure that the plants are as safe as they can be in order to prevent accidents.

The fission process that nuclear power plants use creates nuclear waste which is extremely toxic and harmful to humans. There are specific rules and regulations in place for storing nuclear waste so that the toxins are not emitted into the air and cause danger for the people living near a nuclear plant. The fission process must be designed properly and monitored while in use so that the waste is contained safely if something were to go wrong [5]. Licensing is crucial in making sure every nuclear power plant is running properly and up to date with the regulations.

Children are more susceptible to the radiation emitted from power plants than adults. If a child and an adult are exposed to the same amount of radionuclides, the effect will be greater on the child due to his or her smaller size. Children's cells are also changing at a faster pace and undergoing more divisions because they are in the growth process. A child's developing tissue can be harmed by exposure to radiation. Therefore, we will be looking at childhood leukemia rates as we expect those to be the best indicator of the effect of nuclear power plants [9]. Various types of leukemia will be focused on in our study.

Due to the increased number of leukemia diagnoses in children over the past several years, other studies have begun to look at the effects nuclear power plants may have on these numbers. In particular, a German study known as KiKK, looked at the leukemia rates in children under the age of five near 16 nuclear plant sites from 1980-2003. The study found a 2.2-fold increased cancer rate in children living 5 km away from these sites. The study has been accepted by the German government, which has concluded that there is an increased risk associated with living in close proximity to a nuclear power plant [4].

However, not all studies have shown such a connection. A study conducted by the National Cancer Institute (NCI) concluded that nuclear power plants pose no additional risk to those living near them. This study looked at deaths from 16 different types of cancer in 107 counties in the United States which contain or are closely located to nuclear power plants. Each county was compared to three other counties not containing power plants that were considered to be socioeconomically similar [11].

While some rates increased others decreased, and no clear connection could be made to the power plants. The study calculated the relative risk of developing leukemia before and after the plants became operational. For most cancers, these risk numbers decreased after operation began [11]. The NCI study was designed to examine cancer rates in general, and while childhood leukemia was looked at individually, our study will focus on these cancers solely, as it provides the most beneficial data. The NCI study also compared selected counties containing power plants to a number of control counties. Our study will look at all counties in the continental United States which contain a power plant and will look at their change in mortality rates to determine if nuclear power plants are posing a threat.

Members of the department of Biology, Biostatistics, and Epidemiology at the Medical University of South Carolina conducted a meta-analysis in 2007. The study found an increase of childhood leukemia near power plants but was unable to confirm that this result was directly correlated with the radiation

emitted from the power plants themselves. The study pointed out that a population density bias may exist, meaning that results from highly populated areas will overshadow those areas of lesser population [1]. Our study will be using mortality rates instead of counts in an effort to correct this problem.

II) METHODS

1) DATA

In order to examine whether or not a correlation exists between cancer rates and the distance lived from a nuclear power plant, we needed data covering a variety of variables. These variables were then compared to determine which one played the most significant role in the change in cancer rates. Data from the Centers for Disease Control and Prevention (CDC) provided the number of leukemia related childhood deaths (transform) in 1988 per county in the United States.

Our analysis is based upon mortality rather than the number of incidences reported. It is possible that one case of leukemia can be recorded multiple times, whereas mortality rates are much more accurate. Due to this unreliability of incidence data, we chose to use the mortality data associated with childhood leukemia in each county. We converted the mortality counts provided by the CDC to rates using the county's population for those under 18 (popTotalUnder18 and %TotalUnder18) according to the Bureau of the Census.

We were unable to obtain data on Lovington County in Texas and Yellowstone Park County in Montana, so these counties were omitted from our study. Los Angeles County, California was also omitted due to its heterogeneous population. The county was an outlier for multiple variables and tends to complicate many spatial studies [17].

Using the information provided by the CDC, we separated the data for population into categories based on race and sex [2]. This was done because cancer rates are known to vary amongst males and females as well as various races [8]. We used three different race categories: Caucasians (pWhMaleU20,

pWhFemaleU20), African Americans (pBlMaleU20, pBlFemaleU20), and Other (pOtMaleU20, pOtFemaleU20). We defined other as being all races not included in the Caucasian or African American categories. Data were only used for those under the age of 20 since this age range was more appropriate for our study.

All data were based upon the year 1988 as it is the most recent data that incorporates the dates that the nuclear power plants we are considering were active. It is much more difficult to acquire data from recent years due to increasingly strict privacy laws. Data from 1988, however, are readily available. The U.S. Census Bureau's poverty estimates from 1989 (popUnder18InPov, %Under18InPov) were used due to the lack of availability of 1988 data [14]. We are assuming that these estimates are relatively similar. This data were also only available for those under the age of 18; however, we made the assumption that the rate of poverty would be similar for the rest of our age group and adjusted the data accordingly.

Data on the location of nuclear power plants in 1988 across the continental United States were used from information collected by the Nuclear Regulatory Commission (NRC). Plants located in Alaska and Hawaii were omitted from our study since the data from these states could not be used to look at spatial clustering. This data provided the location of all power plants (#OfNucPlants) using their latitude and longitude coordinates. We reverse geocoded the location of each plant to include the Federal Information Processing Standard (FIPS) code in order to make the location identifiable by county. The NRC data also included the number of days the plant had been operating (avgDaysOperating). We used the starting date of operation until the middle of the year in 1988 to determine our average number of days operating variable. If multiple plants were located within the same county, the average number of operating days was used [15].

2) SPATIAL ANALYSIS

Spatial analysis looks for patterns based on location and measures how similar each location is to its surroundings [19]. Statistical methods for testing spatial clustering can be classified into two groups: methods for determining global clustering and methods for finding a local cluster [19].

In our study we use the population density adjusted exponential weight function in our calculations for global clustering. Define w_{ij} as the i^{th} and j^{th} element in the weight function. The weight function is defined as:

$$w_{ij} = e^{\left(-d_{ij}/k\right)}$$

where d_{ij} is the Euclidean distance between i and j. A parameter, k, increases the sensitivity of the test to large or small clusters, corresponding to large or small values of k. For our study we used values of 0.005, 0.1, 0.5, 1, 20, 50, and 100 for k.

In global clustering, we use various indices to examine an area as a whole and determine if there is any relationship between adjacent locations. For example, we looked at the United States as a whole to see if there were any similarities amongst neighboring counties. Our null hypothesis in performing these tests was that the data are not spatially correlated. The alternative hypothesis was that the data are spatially correlated. We used three of the most common methods to evaluate global clustering in our study: Tango's Index, Moran's I, and Oden's I^*pop. Values near zero indicate there is no spatial clustering and values larger than zero show spatial clustering. It was important to use multiple indices and compare their values since there is no generally accepted method.

The first index we used was Moran's I which is defined as

$$I = \left(\frac{1}{S^2}\right)\frac{\sum_{i}^{N}\sum_{j}^{N} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i}^{N}\sum_{j}^{N} w_{ij}}$$

where y_i is the rate of death in geographic unit *i* and *N* is the number of geographic units. Normally, the number of cases would be used, but since the population varies greatly amongst counties, rates of death were used instead. We defined the mean and variance of the number of cases as $\bar{y} = \frac{1}{N}\sum_i y_i$ and $S^2 = \frac{1}{N}\sum_i (y_i - \bar{y})^2$, respectively. To calculate the *p*-value, we standardized the index as

$$z = [I - E(I)] / \sqrt{Var(I)}$$

so *z* will have a standard normal distribution. We define $E(I) = -1/(N-1)$ and $Var(I) = (N^2 S_1 - NS_2 + 3S_0^2) / [(N-1)(N+1)S_0^2] - (1/N-1)^2$ with $S_0 = \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}$, $S_1 = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(w_{ij} + w_{ji})^2$, and $S_2 = \sum_{i=1}^{N}(w_{i+} + w_{+j})^2$, with $w_{i+} = \sum_{j=1}^{N} w_{ij}$ and $w_{+i} = \sum_{j=1}^{N} w_{ji}$ [10].

Tango's Index was also considered. Aspects of this test allow us to measure goodness of fit which makes this a satisfactory test for determining spatial clustering. Unlike Moran's *I*, Tango's Index takes into account the county's population. Therefore y_i is the number of cases rather than a rate.

Tango's Index was calculated as

$$T = \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}\left(\frac{y_i}{y_+} - \frac{n_i}{n_+}\right)\left(\frac{y_j}{y_+} - \frac{n_j}{n_+}\right)$$

where y_+ = sum of y_i for y_i = number of deaths in geographic unit *i* and n_+ = sum of n_i for n_i = population under 20 at risk for geographic unit *i*.

To find the *p*-value for Tango's Index, we applied a chi-squared approximation. Let *W* be a weight matrix and $p = \left(\frac{n_1}{n_+}, \frac{n_2}{n_+}, \ldots, \frac{n_N}{n_+}\right)$, a vector of proportions. The standardized Tango's Index is represented by

$$T^* = \frac{T - E(T)}{\sqrt{V(T)}},$$

where $E(T) = \frac{1}{y_+} trace(W * V_P)$ and $V(T) = \frac{2}{y_+^2} trace\left((W * V_p)^2\right)$ represent the mean and variance of Tango's index, respectively, with $V_p = diag(p) - p * p'$ where *diag(p)* denotes the diagonal matrix of the elements in *p*. Define $skT^2 = 2\sqrt{2}\, trace\left((W * V_p)^3\right) \Big/ \left(trace\left((W * V_p)^2\right)\right)^{1.5}$ and $dfT = \frac{8}{skT}$. A chi-squared approximation for Tango's Index is $dfT + Tstar\sqrt{2 * skT} \sim \chi^2(dfT)$ [13].

The final index we calculated was Oden's I^*pop. This index is derived from Moran's *I* but integrates the population's size of the counties. We calculated Oden's I^*pop as

$$I^*_{pop} = \frac{n^2 \sum_i \sum_j w^*_{ij}(e_i - v_i)(e_j - v_j) - n(1 - 2\bar{b}) \sum_i w^*_{ii} e_i - n\bar{b} \sum_i w^*_{ii} v_i}{\bar{b}(1 - \bar{b}) \left(y_+^{\,2} + \sum_i \sum_j v_i v_j w^*_{ij} - y_+ \sum_i v_i w^*_{ii} \right)}$$

where $\bar{b} = y_+ / n_+$, $v_i = n_i / n_+$, $e_i = y_i / y_+$, and $w^*_{ij} = w_{ij} \Big/ \sqrt{v_i v_j}$ [12].

In addition to these indices which measure global clustering, we used Local Moran's *I* which is an index of spatial correlation which detects outlying counties. It is calculated for each region using

$$I_{r_i} = \frac{\sum_{j=1}^{N} w_{ij} \frac{y_i - rn_i}{\sqrt{rn_i}} \frac{y_j - rn_j}{\sqrt{rn_j}}}{\sum_{j=1}^{N} w_{ij}},$$

where $r = \sum_{i=1}^{N} y_i \Big/ \sum_{i=1}^{N} n_i$ is the overall rate [6].

3) STATISTICAL ANALYSIS

In order to create a general linear model, we studied the correlation between the variables during our preliminary analysis. We let the childhood leukemia mortality rate be our dependent variable while the variables relating to race, sex, poverty, number of nuclear power plants, and the average number of days operating were independent variables. We created a correlation matrix to determine which independent variables were correlated with each other in order to minimize the effects of multicollinearity and which of the independent variables were most highly correlated with the dependent variable.

To create the best model possible, we used a technique known as backwards elimination. We ran the regression using all of the variables and examined the *p*-value given for each one. If this value was above 0.05, that variable was eliminated and the regression was re-run. This process continued until all *p*-values were in the desired range.

III) RESULTS

1) PRELIMINARY ANALYSIS

In order to use standard regression techniques, we needed to examine our data and perform several tests for normality. Histograms of the data revealed that our dependent variables did not have a normal distribution. A scatter plot of the residuals from our general linear model showed a pattern which also verified that our data were not normally distributed. There were a large number of zeros for the number of deaths which made the data difficult to work with. In order to get the data in a usable format, we attempted to use the Freedman-Tukey transformation but this did not make our data appear normal. We were able to transform our data using the following transformation, $transform = \ln[1000 * (deaths + 1500) / popUnder20]$. This alteration transformed the response variable to give it an approximate normal distribution.

We looked at a correlation matrix and scatter plots of all the variables to determine which variables were most appropriate for our model. These revealed that a few of the variables were highly correlated with each other, and therefore, did not all need to be present in our model. In general, men and women of each race were highly correlated with each other. Since men are more likely to develop leukemia, we used only the data for men [8]. Men were also more correlated with the response variable than women, making them more useful to the model. The following table shows the correlation between all of the variables on the top half of the diagonal. The bottom half of the table below the diagonal gives the *p*-value for the test of the null hypothesis which was that the correlation is zero. The alternative hypothesis was that the correlation is not zero. These *p*-values were computed using Pearson's correlation test.

	transform	pWhMaleU20	pWhFemaleU20	pBlMaleU20	pBlFemaleU20	pOtMaleU20	pOtFemaleU20	popTotalInPov	%TotalInPov	popUnder18InPov	%Under18InPov	#OfNucPlants	avgDaysOperating
transform	1	0.165	0.127	-0.151	-0.149	-0.008	-0.007	-0.406	0.206	-0.383	0.175	-0.107	-0.091
pWhMaleU20	< 0.001	1	0.975	-0.91	-0.912	-0.295	-0.295	-0.176	-0.511	-0.173	-0.471	-0.019	-0.016
pWhFemaleU20	< 0.001	< 0.001	1	-0.91	-0.908	-0.299	-0.297	-0.165	-0.513	-0.163	-0.476	-0.021	-0.016
pBlMaleU20	< 0.001	< 0.001	< 0.001	1	0.997	-0.108	-0.109	0.158	0.447	0.154	0.393	0.029	0.025
pBlFemaleU20	< 0.001	< 0.001	< 0.001	< 0.001	1	-0.108	-0.109	0.159	0.447	0.155	0.393	0.029	0.025
pOtMaleU20	0.673	< 0.001	< 0.001	< 0.001	< 0.001	1	0.997	0.048	0.213	0.051	0.245	-0.018	-0.021
pOtFemaleU20	0.68	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	1	0.047	0.212	0.05	0.244	-0.019	-0.02
popTotalInPov	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.007	0.009	1	0.017	0.993	0.055	0.088	0.061
%TotalInPov	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.349	1	0.036	0.954	-0.067	-0.057
popUnder18InPov	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.004	0.005	< 0.001	0.044	1	0.078	0.078	0.051
%Under18InPov	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.002	< 0.001	< 0.001	1	-0.066	-0.061
#OfNucPlants	< 0.001	0.283	0.24	0.108	0.107	0.303	0.297	< 0.001	< 0.001	< 0.001	< 0.001	1	0.776
avgDaysOperating	< 0.001	0.384	0.367	0.159	0.161	0.25	0.263	0.001	0.001	0.004	< 0.001	< 0.001	1

Table 1. Correlation table

2) SPATIAL STATISTICS

To determine if our data were spatially correlated we analyzed our calculations of Moran's *I*, Tango's Index, and Oden's *I*pop*. When calculating these indices, we used a weight function with varying kappa values. The results from each of these computations were very similar. We determined that there was no statistical significance between them and decided that a kappa value of 0.5 was appropriate for our analysis.

Of the three global indices, Moran's *I* was the only index that indicated a correlation existed and yielded a *p*-value less than 0.001. The *p*-value calculated using Tango's Index, however, was 0.438 which implies that the data were not spatially correlated. The various indices provided by Oden's I*pop were all near zero. This would lead to an insignificant *p*-value.

Using Local Moran's *I*, we determined that there are no outlying counties. The results of this test were insignificant for each county with all indices being near zero, leading us to believe that there are no significant local clusters.

3) REGRESSION

Our final model was created using the following variables with their corresponding values. All of these variables gave a *p*-value less than 0.05 and therefore showed significance.

Variable	**Estimate**	**Standard Error**
Intercept	-3.678	1.054
Percent Caucasian Male Under 20	15.822	2.044
Percent African American Male Under 20	10.896	2.114
Percent Other Male Under 20	10.367	2.152

Total Percentage in Poverty	0.076	0.004
Number of Nuclear Plants	-0.453	0.09

Table 2. Coefficients for general linear model

All of these variables proved to be significant with *p*-values under 0.001. The number of nuclear power plants yielded a negative coefficient. This implies that the number of plants and a counties mortality rate are negatively correlated. All of the other coefficients are positive, meaning that there is a positive correlation between all of the variables and the mortality rates.

One of the assumptions that insures the accuracy of the model is that the residuals are approximately normal. We examined the Q-Q plot of this model's residuals and observed that they fell in a relatively straight line. This is close enough to normal for us to believe that our model is accurate. We also plotted the residuals versus the predicted values to determine if there were any systematic patterns or evidence of non-constant variance. The plot showed that the assumptions were met.

4) GEOGRAPHICAL ANALYSIS

The graph was produced using ArcGIS mapping software. We mapped the death rates per 100,000 people in each county of the continental USA in 1989 with Los Angeles County removed and the location of the nuclear plants active in 1989. This figure uses a technique known as smoothing. Smoothing the data sums the death rates of the county and its neighbors and then divides them by the number of its neighbors to produce a weighted average for that county [7].

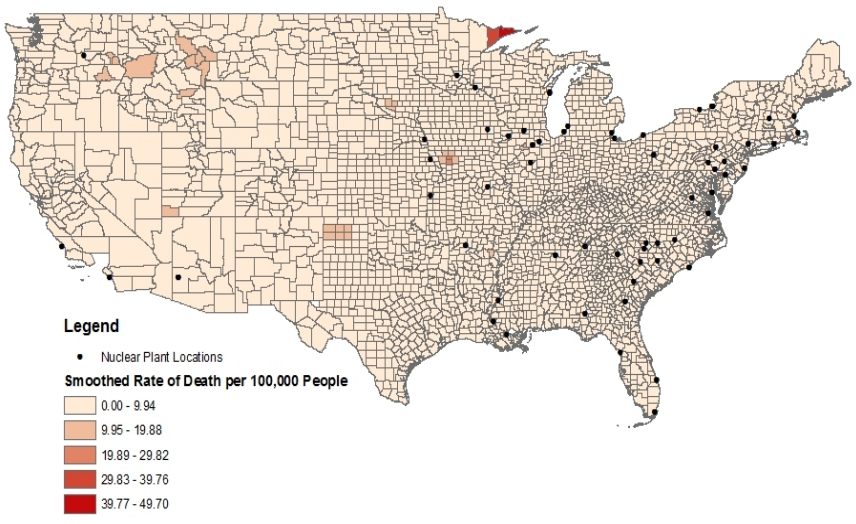

Figure 1. The smoothed rates of death per 100,000 people for each U.S. county in 1989

From this map we can begin to draw conclusions. It is obvious that there are no global patterns. This graph confirms that no correlation between power plants and mortality rates exists. There appear to be outliers in Minnesota, Missouri, Texas, Montana, and on the Utah-Arizona border. Further tests were carried out to deny these conclusions.

IV) DISCUSSION

At the beginning of this study, we expected to find that our data were spatially correlated. We ran multiple tests to check for global clustering. Two of the three indices concluded that no spatial relationship existed. There are more nuclear power plants located in the eastern half of the United States, thus we were anticipating some sort of pattern in this section of the country. However, since our data did not reveal such a correlation, we opted for a linear model rather than using spatial regression techniques.

Our general linear model provided a negative coefficient for the variable representing the number of nuclear power plants. This indicates that there is an inverse relationship between childhood leukemia mortality rates and the number of nuclear power plants in each county which led us to conclude that living in the vicinity of a nuclear power plant does not increase a person's likelihood of developing cancer. Our model also provided a higher coefficient for Caucasians than African Americans. This is in agreement with the known statistic that Caucasians are more likely to develop cancer than other races [8].

Our findings matched those of the National Cancer Institute's which looked at 107 counties and the 62 nuclear sites in or near those counties. This study compared those counties to 292 other counties which did not contain a nuclear plant. The NCI looked at many different cancers and found that there

was a large range between the rates depending upon the type of cancer. They concluded that nuclear power plants could not be blamed for any increase in rates [11].

However, many European studies contrast our findings. The previously mentioned KiKK study which was accepted by the German government claims that living within five kilometers of a power plant increases the risk of cancer [4]. Another study done in France which used very similar methods to our study found a cluster of childhood leukemia cases near the La Hague power plant in Normandy [16].

A study conducted in Sweden in 1995 looked for clusters throughout their study area and then tested the significance of those areas. They specifically checked for higher incidence levels near power plant locations. The study concluded that none of the detected clusters were significant [18].

A study done in the United States reported that an increased risk does exist for those living near a nuclear power plant. The area associated with this risk is much larger than the KiKK study's proposed area due to airborne radiation and the chance of digesting particles found in various food sources. This study looked primarily at incidence rates which would explain their differing conclusions. When they examined mortality rates solely, they did not find any significance indicating that power plants posed a threat [9]. We did not find incidence data to be reliable enough for our study and chose, instead, to use mortality rates.

Several problems occurred while trying to fit our data to a model. Low counts in several counties made various statistical methods difficult to use. For example, many counties reported zero leukemia related deaths which made the data very skewed. Some of the techniques that would normally be applied are not robust enough to handle this problem.

We were also unable to obtain current mortality data due to increased regulations. The fact that we have old data may influence the significance of our results. We also failed to account for other types

of cancer and used only mortality rates associated with childhood leukemia. It is possible that other cancers are impacted by the radiation emitted by nuclear power plants. Our model only takes into account a small number of variables. It is not known what directly causes leukemia; so, we included variables that we assumed to be influential. There are many other factors that may or may not be responsible for the current increase in cancer rates. Not all of these are easily measured or obtainable.

Our given data made it difficult to quantify the various types of power plants. Since only those that emit radiation are necessary for our study, we only wanted to include those that may actually be harming the surrounding inhabitants.

In order to do a more conclusive study, we would like to run similar techniques on more recent data. Current mortality rates would make the study more reliable, but we would also like to obtain more accurate incidence data to better account for all cases of the disease, not just the fatal ones. It would also be beneficial to look at a wider variety of variables and try to create a model that better explains a counties death rate.

In order to check the validity of our model, we would like to perform the same procedures using European data. Our study is in agreement with the NCI's study which looked at similar locations. We would like to perform our analysis on a dataset in which a study has found nuclear power plants to be a cause of leukemia to see if we obtain comparable results.

Works Cited

1. Baker, P., & Hoel, D. (2007). Meta-analysis of standardized incidence and mortality rates of childhood leukaemia in proximity to nuclear facilities. *European Journal of Cancer Care* .

2. Centers for Disease Control and Prevention. (2011, May 16). *Compressed Mortality File*. Retrieved June 29, 2011, from http://www.cdc.gov/nchs/data_access/cmf.htm

3. Feinstein, J. S. (1989). The Safety Regulation of U.S. Nuclear Power Plants: Violations, Inspections, and Abnormal Occurences. *The Journal of Political Economy* , 115-154.

4. Ghirga, G. (2010). Risk of cancer in children living near nuclear power plants. *Italian Journal of Pediatrics* .

5. Green, H. P. (1973). Public Participation in Nuclear Power Plant Licensing: The Great Delusion. *William and Mary Law Review* , 503-504.

6. Jackson, M. C., Huang, L., Luo, J., Hachey, M., & Feuer, E. (2009). Comparison of tests for spatial herogeneity on data with global clustering patterns and outliers. *International Journal of Health Geographics* , 2-3.

7. Lance, W., & Gotway, C. (2004). *Applied Spatial Statistics for Public Health Data.* Hoboken: John Wiley & Sons, Inc.

8. Lightfoot, T. (2005). Aetiology of Childhood Leukemia. *Bioelectromagnetics Supplement* , S5-S11.

9. Mangano, J. J., Sherman, J., Chang, C., Dave, A., Feinberg, E., & Frimer, M. (2003). Elevated Childhood Cancer Incidence Proximate to U.S. Nuclear Power Plants. *Archives of Environmental Health* , 74.

10. Moran, P. (1950). Notes on continuous stochastic phenomena. *Biometrika* , 17-23.

11. National Cancer Institute. (2011, April 19). *No Excess Mortality Risk Found in Counties with Nuclear Facilities*. Retrieved June 28, 2011, from http://www.cancer.gov/cancertopics/factsheet/Risk/nuclear-facilities

12. Oden, N. (1995). Adjusting Moran's I for population density. *Stat Med* , 17-26.

13. Tango, T. (1995). A class of tests for detecting 'general' and 'focused' clustering of rare diseases. *Stat Med* , 2323-34.

14. U.S. Census Bureau. (2010, December 8). *Small Area Income and Poverty Estimates*. Retrieved June 29, 2011, from http://www.census.gov/did/www/saipe/data/statecounty/data/index.html

15. United States Government. (2009, August). *U.S. Commercial Nuclear Power Reactors*. Retrieved June 29, 2011, from data.gov: http://explore.data.gov/Energy-and-Utilities/U-S-Commercial-Nuclear-Power-Reactors/rfjn-7hyh

16. Viel, J.-F., Pobel, D., & Carrre, A. (1995). Incidence of leukaemia in young people around the La hague nuclear waste reprocessing plant: A sensitivity analysis. *Statistics in Medicine* , 2459-2472.

17. Waller, L. A., Hill, E. G., & Rudd, R. A. (2006). The geography of power: Statistical performance of tests of clusters and clustering in heterogeneous populations. *Statistics in Medicine* , 853-865.

18. Waller, L. A., Turnbull, B. W., Gustafsson, G., Hjalmars, U., & Andersson, B. (1995). Detection and assessment of clusters of disease: An application to nuclear power plant facilities and childhood leukaemia in Sweden. *Statistics in Medicine* , 3-16.

19. Waller, L., & Gotway, C. A. (2004). *Applied Statistics for Public Health.* New York: Wiley.