A Spatial Study of the Location of Superfund Sites and Associated Cancer Risk

Raid Amin, Arlene Nelson, and Shannon McDougall
Department of Mathematics and Statistics, University of West Florida, Pensacola, FL

ABSTRACT

Superfund sites are geographic locations selected by the U.S. Environmental Protection Agency as having extreme toxic chemical spills. In this article, we address three main research questions: (1) Are there geographical areas where the number (or density) of Superfund sites is significantly higher than in the rest of the USA? (2) Is there an association between cancer incidence and the number (or density) of Superfund sites? (3) Do counties with Superfund sites have higher proportions of minority populations than the rest of the USA? We study the geographic distribution of the overall cancer incidence rate (2007–2011) in addition to the geographic variation of Superfund sites for 2013. We used the disease surveillance software package SaTScan with its scan statistic to identify locations and relative risks of spatial clusters in cancer rates and in Superfund site count and density. We also used the surveillance software FlexScan to support and complement the results obtained with SaTScan. We find that geographic areas with Superfund sites tend to have elevated cancer risk, and also elevated proportions of minority populations.

ARTICLE HISTORY
Received December 2016
Accepted November 2017

KEYWORDS
Cancer; Environmental justice; EPA; Pollution; Scan statistic; Superfund site

1. Introduction

Chemical spills can cause severe environmental and health problems (Kearney 2008; Russi, Borak, and Cullen 2008; Goodman, Hudson, and Monteiro 2010; Boberg, Lessner, and Carpenter 2011; Lu, Lessner, and Carpenter 2014; EPA April 2015). The Environmental Protection Agency (EPA) has a special program to address the clean-up of some of the most problematic spill sites in the country. Hazardous waste sites that meet certain criteria are designated as Superfund sites and are placed on the National Priorities List (NPL), which is a list of sites eligible for long-term cleanup. Some common contaminants found at Superfund sites include arsenic, lead, mercury, and polychlorinated biphenyls (PCB, EPA April 2015). These toxins, along with others, can seep into surface water, groundwater, soil, air, and even buildings. One study found associations between excess cancer deaths from 1970–1979 and hazardous waste site locations that contaminated the drinking water (Griffith et al. 1989). The same study also found a cluster of elevated rates of gastrointestinal cancers in counties located in EPA Region 3 (Delaware, Maryland, Pennsylvania, Virginia, and West Virginia), a region with many Superfund sites.

There have been several environmental justice studies conducted to determine which segments of the population are most adversely affected by Superfund sites (Boer et al. 1997; Stretesky and Hogan 1998; Burwell-Naney et al. 2013). Most research suggests that non-white populations as well as Hispanic populations are more likely to live near Superfund sites. The same studies also found that areas with higher levels of poverty and lower levels of education may also be impacted. However, race and ethnicity seem to play a larger role than poverty and education (Boer et al. 1997; Burwell-Naney et al. 2013). Recently, potential associations between Superfund locations in Florida and cancer incidence were studied (Kirpich and Leary 2016), but there is no published study on clusters of superfund sites covering the entire contiguous USA. This study uses modern disease surveillance software packages to identify clusters of counties with Superfund sites and with associated cancer rates in the contiguous USA. A map of EPA NPL sites in the United States can be found at https://epa.maps.arcgis.com/apps/webappviewer/index.html?id=33ecbdfd1b4c3a8b51d416956c41f1 where yellow marks stand for current NPL sites, while green and red marks stand for deleted NPL sites and proposed NPL sites, respectively.

1.1. Statement of Problem

This article studies the (not necessarily causal) association between the location of Superfund sites and the corresponding cancer rates in all 48 contiguous states in the USA. A limitation of this study is that the results do not take into account that some people in some counties may have moved to/from other counties during or before the years covered in this project. It should also be noted that lifestyle (such as smoking and physical activity) may impact cancer rates, and such factors have not been directly included in this study. The National Cancer Institute states that the NCI conducts and funds research to “identify and evaluate a range of exposures and risk factors that may be associated with cancer.” This includes among several factors “diet and nutrition, tobacco use, alcohol use, energy balance, physical activity, and...”

CONTACT
Raid Amin  ramin@uwf.edu  Department of Mathematics and Statistics, University of West Florida, 11000 University Parkway, Pensacola, Fl 32514.
obesity” in addition to “air pollutants, water pollutants, and chemicals” (NIH National Cancer Institute October 2017).

The EPA defines Environmental Justice as “the fair treatment and meaningful involvement of all people regardless of race, color, national origin, or income, with respect to the development, implementation, and enforcement of environmental laws, regulations, and policies” (EPA October 2017). Ash and Fetter (2004) studied the social and economic factors as they correlate with air pollution exposure. They concluded that “blacks tend to live both in more polluted cities in the U.S. and in more polluted neighborhoods within cities. Hispanics live in less polluted cities on average, but they live in more polluted areas within cities” (Ash and Fetter 2004). Anderton et al. (1994) stated that “facilities for treatment, storage, and disposal of hazardous wastes are located disproportionately in minority areas. This study also addresses the question of whether minorities disproportionately live in polluted areas, but more specifically, whether they live in counties containing Superfund sites.

Specifically, we aim to answer the following questions:
1. Are there geographical areas where the number (or density) of Superfund sites is significantly higher than in the rest of the contiguous USA?
2. Is there an association between the number (or density) of Superfund sites and corresponding cancer rates?
3. Are counties with Superfund sites more likely to have higher rates of minority populations than the rest of the contiguous USA?

It is important for county, state, and even the national levels of government to be aware of potential health threats to people who live near Superfund sites. Counties may need additional resources to ensure they can protect their residents from hazards associated with pollution. Furthermore, people have an implicit right to know whether their cancer risk is elevated due to environmental factors. If local governments know that a particular site is associated with increased cancer risk, then steps to remediate the site should be prioritized.

1.3. Limitations

Our study aggregated all data at the county level. Since not all Superfund sites affect large areas, some localized effects will not be found using data at this scale. Census tract or zip code tabulation area data may be more effective in finding impacts at a smaller geographic level.

The cancer information from the NCI was reported as an average of 5 years of counts. While this should work well for review, it is possible that 1 year may have had an unusually high or low rate for one particular county. Such spikes will not be apparent in this data. Furthermore, the NCI states in its data files that the individual state cancer registries may have more local or current data (Center for Disease Control: National Cancer Institute December 2016). Since gathering data from each state would have been a formidable task, we accept the limitation that our results are only as accurate as the data held by the NCI.

2. Methods

2.1. Spatial Analysis

We used two spatial analysis tools, SaTScan™ v9.4.2 and FlexScan v3.1.2. SaTScan™ is a disease surveillance software for spatial and temporal analysis, developed by Martin Kulldorff, and it is widely used to study disease epidemics (Kulldorff, 2015). SaTScan™ is able to handle many different models, including the Poisson, Bernoulli, and Normal. The spatial analysis in SaTScan™ allows the use of circular or elliptical windows with its scan statistics (Kulldorff 2007). We chose for SaTScan to use circular windows. For each location being analyzed, the software scans circular windows ranging in size from one county to the upper limit defined by the user, noting the number of observed cases and the expected observations inside each window at each location. The “most likely cluster” is defined as the cluster with the maximum likelihood, and it is the identified cluster that is least likely to be due to chance. In
the Poisson model, the likelihood function is proportional to
\[
\left( \frac{E^n}{n^n} \left( \frac{N - n}{N - E} \right)^{N-n} \right) I(n > E),
\]
where \( n \) is the number of counts within the scan window, \( N \) is the total number of counts in the contiguous USA, and \( E \) is the expected number of counts under the null hypothesis of equal risk (of cancer) for all counties within the scan window. In the normal distribution model with SaTScan, the likelihood under the null hypothesis is given by
\[
L_0 = \prod \frac{e^{-\mu_i} \mu_i^{x_i}}{\sqrt{2\pi \sigma^2}},
\]
where \( \mu = X/N \) and \( \sigma^2 = \sum_{i=1}^{N} (x_i - \mu)^2 / N \) are the maximum likelihood estimates of the mean and variance, respectively. \( X \) is the sum of the observed continuous values \( x_i \). Each continuous observation is at a spatial location with latitude and longitude coordinates. The log likelihood is used to calculate the maximum likelihood estimators for the mean inside the scan circle and the mean outside the circle (Kulldorff 2007, 2015). Detailed information on the normal model is at www.satscan.org.

The software FlexScan uses a flexible spatial scan statistic developed by Kunihiko Takahashi and Toshiro Tango, which finds smoothly shaped clusters, rather than only circular or elliptical clusters, by examining geographical areas adjacent to each other (Tango and Takahashi 2012). Unlike SaTScan™, FlexScan is limited to spatial analysis only (i.e., not temporal) and it does not allow for a cluster analysis when the response variable is continuous (Takahashi, Yokoyama, and Tango 2015).

We propose a Superfund density metric as the ratio of the number of superfund sites per county to the area of the county. This metric is continuous in nature, so we used the normal model in SaTScan™ to search for clusters of high Superfund density. The normal probability model in SaTScan™ has a null hypothesis that all observations come from the same cluster, while the alternative hypothesis states that at least one cluster has a higher or lower mean than the area outside the cluster. Clusters are identified by examining the likelihood value for each of many overlapping circles (Kulldorff, Huang, and Konty 2009). Superfund density is available for all 3109 counties in the 48 contiguous states. The Superfund density was first normalized to a standard normal distribution in SAS v9.4. We then ran SaTScan™ using a case file containing the county code and Superfund density for each county. The location file was a file of county codes and the coordinates of the county centroids. We used a purely spatial analysis which identified circular-shaped clusters, using a maximum cluster size of 5% of the population. This maximum size imposed an upper limit on cluster size.

We examined cancer incidence data using a Poisson regression in SAS v9.4 to age-adjust the raw cancer counts. We then ran the Poisson model in FlexScan using both the restricted likelihood ratio and the flexible scan statistic. We used alpha = 0.2 for the restricted likelihood ratio and 15 for the maximum cluster size, which are both defaults in the software. The analysis for cancer incidence clusters was run for 2900 counties because cancer incidence rates were not available for three states. We adjusted the FlexScan county adjacency file to remove all references to counties in these three states so the cluster analysis would run correctly. The software treated the missing states in the same way it would treat any other separation between counties, such as a large lake. To examine the effects of the Superfund density on cancer incidence rates, we ran another Poisson regression adding Superfund density to the model. As before, we ran the Poisson model in FlexScan for the cluster analysis.

2.2. Other Statistical Tests

We used the Jonckheere-Terpstra (JT) test (Jonckheere 1954) in SAS to study the relationship between the number of Superfund sites in a county and the cancer incidence rate of that county. The JT test is a nonparametric test for ordered differences among groups. We used this test to test the following hypothesis:
\[
H_0 : \mu_1 = \mu_2 = \cdots = \mu_k \text{ against the alternative } H_1 : \mu_1 \leq \mu_2 \leq \cdots \leq \mu_k \text{ where at least one population median is strictly less than at least one other population median.}
\]

\[
J = \frac{\sum n^2 - \sum n_i^2}{4} - \frac{\sum n_i^2 (2n_i + 3)}{72},
\]

where \( n_{ab} \) is the number of observations in group \( b \) that are greater than each observation in group \( a \), \( N \) is the total sample size, \( N = \sum_{i=1}^{k} n_i \), and where \( n_i \) is the number of observations in sample \( f \). The \( J \) values are compared with values from a standard normal distribution.

Superfund sites were measured by placing each county into one of five categories. Category one contained counties without any Superfund sites, categories two, three, and four contained counties with one, two, and three Superfund sites, respectively, while category five contained counties with at least four Superfund sites. Twenty-two counties had 10 or more Superfund sites. The cancer incidence rates were then compared across the five categories. The JT test addresses the question whether cancer rates increase on average as the number of superfund sites per county increases.

To determine the relative effects of several demographic variables on the make-up of a county containing a Superfund site, a stepwise logistic regression was run in SAS v9.4. Of the 2900 counties with available data on cancer incidence rates, 710 contained at least one Superfund site. Counties with a Superfund site were assigned a value of 1, and those without a Superfund site were assigned a zero. The logistic regression was then run to determine the likelihood of a county containing a Superfund site based on the percentage of people in poverty, the percentage of African Americans, the percentage of Hispanics, the percentage of high school drop-outs, the cancer incidence rate, and the percentage of males in the county.

2.3. Assessing the Sites for Inclusion on the National Priorities List

The EPA has set criteria for determining whether an uncontrolled waste site is included in the National Priorities List (NPL). After a preliminary assessment and site inspection, a Hazard Ranking System (HRS) is used for assigning a score, and if a site gets an HRS score of 28.50 or greater, it is considered for inclusion in the NPL. The EPA states that the information
used to assign HRS scores is not sufficient to determine the extent of contamination for a waste site. The EPA identifies four exposure pathways on which scores could be obtained (Gamper-Rabindran, Shanti, and Christopher 2017).

(i) groundwater migration pathway ($S_{GW}$),
(ii) surface water migration pathway ($S_{SW}$),
(iii) soil exposure pathway ($S_s$), and
(iv) air migration pathway ($S_A$).

If data are available on each of the four pathways, the HRS score is obtained as the root mean square average of the four pathway scores:

$$S = \sqrt{\frac{S_{GW}^2 + S_{SW}^2 + S_s^2 + S_A^2}{4}}.$$  

The groundwater migration pathway score is obtained for each aquifer at the site, and the highest score is used in the calculation of $S_{GW}$. As for the surface water migration pathway, the overland flow and the release caused by flooding is used in calculating $S_{SW}$ in addition to other input values. The soil exposure pathway is measured by $S_s$ and it is based on the exposure of the populations near that location. The air migration pathway score $S_A$ is measured from the exposure to chemical releases and particulate releases to the air that the population is exposed to within one mile from the site.

Often the EPA cannot obtain all four pathways scores due to cost issues. In order to be cost-effective, the EPA does not collect additional data to score each pathway. Instead, many sites are scored on only a single pathway. The HRS score in some cases simply reflects the number of pathways scored and the amount of data collected rather than being a measure of the severity of a chemical spill, even though higher HRS scores reflect higher degrees of threat when data are obtained on all four pathways. The EPA sometimes identifies sites with HRS scores exceeding 28.5 but instead of listing them on the NPL, defers them to other state programs. More sites are deferred to state programs than are listed on the NPL (Gamper-Rabindran, Shanti, and Christopher 2017).

In addition to using the HRS scoring system, the EPA also allows each state to designate a State Top Priority designation based on the CDC’s Agency for Toxic Substances and Disease Registry (ATSDR) health advisory criteria. Only one site may be listed by each state as its top priority on the NPL, regardless of its HRS score. There are currently 41 states that have used their top toxic picks. The EPA has listed 14 sites based on ATSDR health advisory criteria regardless of the corresponding HRS score (Gamper-Rabindran, Shanti, and Christopher 2017). It should be noted that the HRS score is used only as a screening tool, so once a decision has been made regarding listing on the NPL (or not), such a decision will not change. It is not advisable to simply sum up HRS scores within each county that contains multiple superfund sites since not all pathways for all sites are scored, and the resulting total HRS score may be meaningless. Also, the age of a chemical spill may differ considerably from its year of designation on the NPL, and it is often unclear when a specific chemical spill may have occurred. While in cancer studies the latency and duration of exposures can impact observed cancer rates, the HRS does not consider such latency or any other epidemiological information beyond toxicity of the carcinogenic chemicals and potential exposure by individuals (Currie, Greenstone, and Moretti 2011).

3. Results

3.1. Superfund Density

Since Superfund density is a continuous variable, the normal model in SaTScan was used. SaTScan found one significant cluster (with $p < 0.05$) of high Superfund density in parts of Delaware, New Jersey, New York, and Pennsylvania. There was also a secondary cluster in Virginia. This cluster is not significant ($p = 0.069$), but since its mean Superfund density value is very high, it is an area that should also be monitored. The map in Figure 1 shows both the significant cluster in yellow, and the secondary cluster in red. The details of the clusters can be found in Table 1. The significant cluster has a low poverty rate (11.37%) compared to the poverty rate in the contiguous USA (16.21%), but its rate of African Americans is quite high (17.77%) relative to the corresponding rate in the contiguous USA (9.52%). The secondary cluster has a poverty rate that is similar to the poverty rate in the contiguous USA, but it has a very high rate of African Americans (49.87%).

3.2. Cancer Incidence and Superfund Sites

To investigate the link between cancer incidence and Superfund sites, we first performed a cluster analysis to identify the areas with higher than expected cancer incidences. We ran a Poisson regression to age adjust the raw cancer incidence counts. This adjustment removed the effects of age from each county’s cancer incidence count, and we used these results in a geographical cluster analysis to find areas where cancer incidence rates are significantly higher than expected. The results from our Poisson model with the restricted likelihood ratio run in FlexScan can be seen in Figure 2. The map shows all geographical clusters of increased cancer incidence. It must be noted that Union County, Florida has a large prison hospital population which could cause it to show up as an outlier. The inmate population of the hospital is not counted in the Union County population at risk, but the cases diagnosed at the hospital are counted in the Union County cancer incidence (Ren, Lim, and Hylton 2012).

The most likely cancer incidence cluster is the cluster with the highest likelihood ratio. It contains counties in Delaware, New Jersey, and Pennsylvania and has a relative risk of 1.32. This means the population within the cluster has a 17% higher risk of cancer than the rest of the U.S. Many of the counties in the most likely cluster are also found in the statistically significant cluster of increased Superfund density. Five of the clusters with the highest cancer incidence rates contained no Superfund sites at all, while the two clusters with the highest likelihood ratios contained significant numbers of Superfund sites. The details of the 10 cancer incidence clusters with the highest relative risk are given in Table 2.

Next, we adjusted the cancer incidence counts by both age and Superfund density and ran a second cluster analysis in FlexScan. Such an adjustment removes from cancer counts effects due to age or due to differences in the number of Superfund sites in the counties. The map shown in Figure 3 shows the
effects of this additional adjustment. The counties in dark red are found in both the age-adjusted and age-Superfund density-adjusted models. The blue counties, listed in Table 3 as “counties removed,” are those counties that were in an age-adjusted cluster, but not an age-Superfund density-adjusted cluster. These are the counties that have cancer incidence rates which can be associated with Superfund density. Finally, the light red counties are those that only appear in the age-Superfund density adjusted results. Table 3 details the age-adjusted cancer incidence clusters that contain at least one county with a high cancer incidence rate attributed to Superfund density.

The Jonckheere-Terpstra (JT) test was used to test whether there is a monotonic increasing trend in cancer incidence as the number of Superfund sites increases. The test results are significant with test statistic \( Z = 8.5341 \) and \( p < 0.0001 \). The results show a significant trend of increasing cancer incidence rates, detailed in Table 4, as the number of Superfund sites in a county increases. The trend can also be seen in the graph shown in Figure 4. Each of the three quartiles (q25%, q50%, q75%) is monotonically increasing with the number of Superfund sites per county.

3.3. County Characteristics

The stepwise logistic regression accepted into the model all the variables listed in Table 5, in addition to the covariate age, and the model had a concordant rate of 72.8%, which is considered high. By using these covariates, the logistic model can correctly categorize 72.8% of the 3109 counties as either having or not having at least one Superfund site. The covariates in the model can be defined as follows: (i) Poverty: Percent county population living below the USA poverty line for a household, where $11,139 is the poverty threshold in 2010 for a one-person household and where the threshold is $14,218 for a two-person

Table 1. High superfund density cluster details.

| Area/cluster states | Contiguous U.S. | DE, NJ, NY, PA | VA |
|---------------------|-----------------|----------------|----|
| Counties            | 29 Counties     | 5.63           | 11.78 |
| Mean SF density inside cluster | – 0.057 | – 0.008 |
| Mean SF density outside cluster | 0.001 | 0.069 |
| p-Value             | 0.37            | 9.06           | 499.99 |
| Average superfund density (per 1000 km²) | 450.90 | 479.95 |
| Average cancer incidence (Excl. KS, MN, NV) | 16.21% | 11.37% |
| Poverty percentage  | 9.52%           | 17.77%         | 49.87% |
| Average African American percentage | 16.21% | 11.37% | 49.87% |
Figure 2. Map of age-adjusted high cancer incidence clusters.

Table 2. Top 10 high cancer incidence clusters.

| States | Counties | Relative risk | Likelihood ratio | p-Value | Average superfund density of counties | Total number of SF sites in cluster | Average poverty percent | Average African American percent |
|--------|----------|---------------|------------------|---------|--------------------------------------|------------------------------------|------------------------|-------------------------------|
| FL     | Union    | 2.65          | 74.35            | 0.001   | 0.00                                 | 0                                  | 18.44%                 | 23.18%                       |
| VA     | Williamburg | 1.89       | 16.10            | 0.004   | 0.00                                 | 0                                  | 18.38%                 | 14.89%                       |
| MS     | Calhoun, Chickasaw, Itawamba, Lee, Monroe, Webster, Yalobusha | 1.89 | 16.78 | 0.003 | 0.00 | 0 | 20.44% | 27.97% |
| KY     | Breathitt, Clay, Estill, Garrard, Jackson, Knox, Laurel, Lincoln, Madison, Powell, Pulaski, Rockcastle, Wolfe | 1.18 | 28.90 | 0.001 | 0.00 | 0 | 28.01% | 1.56% |
| LA, MS | Washington, Copiah, Hinds, Lawrence, Lincoln, Marion, Pike | 1.18 | 29.03 | 0.001 | 0.25 | 3 | 23.86% | 42.66% |
| AL, GA | Russell, Muscogee Kent, Atlantic, Burlington, Camdem, Cape May, Cumberland, Gloucester, Mercer, Monmouth, Ocean, Salem, Delware, Philadelphia | 1.17 | 15.08 | 0.10 | 0.00 | 0 | 20.32% | 45.08% |
|       | * DE, NJ, PA | 1.17 | 411.12 | 0.001 | 6.55 | 12 | 11.53% | 18.25% |
| KY     | Caldwell, Christian, Davless, Hopkins, McLean, Ohio, Webster | 1.15 | 14.56 | 0.012 | 0.42 | 3 | 18.72% | 6.78% |
| IL     | De Witt, Fulton, McLean, Macon, Mason, Peoria, Tazewell, Woodford | 1.15 | 36.59 | 0.001 | 0.00 | 0 | 12.95% | 6.49% |
| MI     | Genese, Ingham, Jackson, Macomb, Shiawassee, Washtenaw, Wayne | 1.15 | 197.63 | 0.001 | 2.71 | 4 | 17.06% | 15.59% |
Table 3. Age-adjusted cancer incidence clusters affected by adjustment for superfund density (p < 0.01).

| General cluster location | Number of initial counties | Relative risk of age-adjusted cancer incidence cluster | Counties removed after adjusting for SF density | Average SF density of removed counties per 1000 km² |
|--------------------------|---------------------------|-----------------------------------------------------|-----------------------------------------------|--------------------------------------------------|
| NY                       | 7                         | 1.10                                                | Orange                                        | 1.84                                             |
| RI                       | 7                         | 1.08                                                | Providence                                    | 7.09                                             |
| NY                       | 12                        | 1.12                                                | Chenango, Cortland                            | 0.60                                             |
| PA                       | 9                         | 1.07                                                | Bucks, Montgomery                             | 9.62                                             |
| IL                       | 9                         | 1.09                                                | Morgan                                        | 0.00                                             |
| WA                       | 8                         | 1.09                                                | Island, Kitsap                                | 3.48                                             |
| FL                       | 6                         | 1.07                                                | Hillsborough                                  | 5.19                                             |
| IL                       | 8                         | 1.15                                                | McLean                                        | 0.00                                             |
| VT                       | 9                         | 1.11                                                | Bennington                                    | 2.28                                             |
| PA                       | 7                         | 1.12                                                | Clinton                                       | 0.43                                             |
| NY                       | 4                         | 1.03                                                | Bronx, Kings, New York                         | 2.66                                             |
| KY/TN                    | 10                        | 1.09                                                | Bell, Harlan, Hamblen, Hawkins                 | 0.21                                             |

Table 4. JT test results for cancer incidence and number of superfund sites.

| Number of superfund sites | Number of counties | Average cancer incidence rate per 100,000 |
|---------------------------|--------------------|------------------------------------------|
| 0                         | 2190               | 446.85                                   |
| 1                         | 415                | 459.20                                   |
| 2                         | 112                | 462.08                                   |
| 3                         | 78                 | 466.74                                   |
| 4+                        | 105                | 478.90                                   |

household (U.S. Census Bureau 2016a), (ii) African American: Percent county population listed as African American based on the Census Bureau (U.S. Census Bureau 2016b), (iii) Hispanic: Percent county population listed as Hispanic based on the Census Bureau (U.S. Census Bureau 2016b), (iv) High School Dropout: Percent of county population over age 25 without a high-school education (U.S. Census Bureau 2016c), (v) Cancer Incidence Rate: A cancer incidence rate is defined by the CDC as the number of new cancers of a specific type occurring in a county population during a year (Center for Disease Control: National Cancer Institute 2016) (such a rate is usually expressed as the number of cancers per 100,000 people at risk), (vi) Male: Percent county population self-identified as being male (U.S. Census Bureau 2016d).

Table 5 gives the maximum likelihood estimates for the logistic regression model. The odds ratios are shown in Table 6, and with these results we can make several observations about counties having at least one Superfund site. They are likely to have a higher rate of minorities, as measured by both race and ethnicity. The population has a higher level of education, as measured by the high school drop-out rate. Interestingly, they are likely to have a lower percentage of males. The US Census Bureau uses the “sex ratio” as a measure to describe the balance between males and females. Geographic regions in the USA have sex ratios that vary, with higher rates of males in Western states and lower rates of males in Northeastern states. There are more
males at younger ages and more females at older ages (U.S. Census Bureau 2011).

When using the full model, cancer incidence and poverty rates did not significantly contribute anything to the model because their odds ratios were so close to one. This is most likely a result of confounded effects (with some of the other covariates in the model).

### 4. Conclusion

We have identified a significant positive association between Superfund density and overall cancer rates across the 48 contiguous USA, in addition to a significant trend for number of superfund sites per county and the corresponding cancer rates. Our results show that geographic areas with greater numbers of Superfund sites tend to have elevated cancer risk across the 48 contiguous states, and such counties also have elevated rates of minority populations.

Obviously, there may be important confounders. From an environmental justice standpoint, minorities may be more likely to live near Superfund sites, for example, if the presence of a site depresses real estate values, then poorer people, often minorities, are likely to rent and own in that area (Ash and Fetter 2004).

The appearance of drop-out rates and gender in the model is more difficult to interpret. There are large regional differences in high school drop-out rates, and this may be conflated with Superfund site location. Regarding gender, aside from countries with prison populations, there is little variation, and this deserves deeper examination on a county by county basis.

### Table 5. Maximum likelihood estimates for the logistic regression model to predict the presence of superfund site(s).

| Parameter            | Estimate | Standard error |
|----------------------|----------|----------------|
| Intercept            | 2.285    | 1.548          |
| pctPoverty           | −0.016   | 0.011          |
| pctAfrican American  | 2.420    | 0.380          |
| pctHispanic          | 4.730    | 0.424          |
| pctHSDrop            | −0.117   | 0.011          |
| CancerIncidence      | 0.008    | 0.001          |
| pctMale              | −11.091  | 2.912          |

### Table 6. Odds ratio estimates from step-wise logistic regression.

| Effect                | Point estimate | 95% Wald confidence limits |
|-----------------------|----------------|----------------------------|
| Poverty               | 0.984          | 0.963                      |
| African American %    | 11.251         | 5.347                      |
| Hispanic %            | 113.236        | 49.308                     |
| High school drop-out %| 0.890          | 0.871                      |
| Cancer incidence rate | 1.008          | 1.006                      |
| Male %                | <0.001         | <0.001                     |

### References

Anderton, D. L., Anderson, A. B., Oakes, J. M., and Fraser, M. R. (1994), “Environmental Equity: The Demographics of Dumping,” Demography, 31, 229–248. doi: 10.2307/2061884. [2]

Ash, M., and Fetter, T. R. (2004), "Who Lives on the Wrong Side of the Environmental Tracks? Evidence from the EPAs Risk-Screening Environmental Indicators Model," Social Science Quarterly, 85, 441–462. doi:10.1111/j.0038-4941.2004.08502011.x. [2,8]

Boberg, E., Lessner, L., and Carpenter, D. O. (2011), “The Role of Residence near Hazardous Waste Sites Containing Bin the Development of Hematologic Cancers in Upstate New York,” International Journal of Occupational Medicine and Environmental Health, 24, 327–338. doi:10.2478/s13382-011-0037-8. [1]

Boer, J., Pastor, M. Jr., Sadd, J. L., and Snyder, L. D. (1997), “Is there Environmental Racism? The Demographics of Hazardous Waste in Los Angeles County,” Social Science Quarterly, 78, 793–810. [1]

Burrwell-Naney, K., Zhang, H., Samantapudi, A., Jiang, C., Dalemarre, L., Rice, L., Williams, E., and Wilson, S. (2013), "Spatial Disparity in the Distribution of Superfund Sites in South Carolina: An Ecological Study," Environmental Health, 12, 96. [1]

Center for Disease Control: National Cancer Institute (2016), "State Cancer Profiles," [online]. Available at https://www.statecancerprofiles.cancer.gov/incidencerates/. [2,7]

Currie, J., Greenstone, M., and Moretti, E. (2011), “Superfund Cleanups and Infant Health,” American Economic Review, 101, 435–441. doi:10.1257/aer.101.3.435. [4]

EPA (2015), “Cleaning Up the Nation’s Hazardous Wastes Sites,” United States Environmental Protection Agency’s Learn the Issues | Science and Technology. [Online]. Available at http://www.epa.gov/superfund/. [1,2]

———. (2017), “Environmental Justice,” United States Environmental Protection Agency. [Online]. Available at https://www.epa.gov/environmentaljustice. [2]
Gamper-Rabindran, T., and Shanti, C. (2017), "Does Cleanup of Hazardous Waste Sites Raise Housing Values? Evidence of Spatially Localized Benefits," *Journal of Environmental Economics and Management*, 65, 345–360. [4]

Goodman, J. E., Hudson, T. C., Monteiro, R. J. M. (2010), "Cancer Cluster Investigation in Residents near a Municipal Landfill," *Human and Ecological Risk Assessment: An International Journal*, 16, 1339–1359. doi:10.1080/10807039.2010.526504. [1]

Griffith, J., Duncan, R., Riggan, W., and Pellom, A. (1989), "Cancer Mortality Among United States Counties with Hazardous Waste Sites and Sole Source Drinking Water Contamination," *Archives of Environmental Health: An International Journal*, 44, 69–74. [1]

Jonckheere, A. R. (1954), "A Distribution-Free k-Sample Test Against Ordered Alternatives," *Biometrika*, 41, 133–145. doi:10.1093/biomet/41.1-2.133. [3]

Kearney, G. (2008), "A Procedure for Detecting Childhood Cancer Clusters Near Hazardous Waste Sites in Florida," *Journal of Environmental Health*, 70, 29–34. [1]

Kirkpich, A., and Leary, E. (2016), "Superfund Locations and Potential Associations with Cancer Incidence in Florida," *Statistics and Public Policy*, 4, 1–9. [1]

Kulldorff, M. (2007), "A Spatial Scan Statistic," *Communications in Statistics-Theory and Methods*, 26, 1481–1496. doi:10.1080/03610920708831995. [2,3]

Kulldorff, M. (2015), "SaTScan User Guide," Available at [http://www.satscan.org/]. [2,3]

Kulldorff, M., Huang, L., and Konty, K. (2009), "A Scan Statistic for Continuous Data Based on the Normal Probability Model," *International Journal of Health Geographics*, 8. [3]

Lu, X., Lessner, L., and Carpenter, D. O. (2014), "Association between Hospital Discharge Rate for Female Breast Cancer and Residence in a Zip Code Containing Hazardous Waste Sites," *Environmental Research*, 134, 375–381. doi:10.1016/j.envres.2014.07.005. [1]

NIH National Cancer Institute (2017), "Why Research on Causes of Cancer is Critical to Progress against the Disease," National Cancer Institute's Research on Causes of Cancer. [Online]. Available at [https://www.cancer.gov/research/areas/causes]. [2]

Ren, C., Lim, S., Hylton, T., Huang, Y., Button, J., Wohler, B. Levin, G., and MacKinnon, J. (2012), *Florida Annual Cancer Report: 2008 Incidence and Mortality*, Tallahassee: Florida Department of Health. [4]

Russi, M. B., Borak, J. B., and Cullen, M. R. (2008), "An Examination of Cancer Epidemiology Studies among Populations Living Close to Toxic Waste Sites," *Environmental Health*, 7, 7–32. [1]

Stretesky, P., and Hogan, M. J. (1998), "An Analysis of Superfund Sites in Florida," *Social Problems*, 45, 268–287. doi:10.2307/3097247. [1]

Surveillance, Epidemiology, and End Results (SEER) Program (2014), "SEER*Stat Database: State Cancer Registry and CDC’s National Program of Cancer Registries January 2014 Sub (1973-2011) <Katrina/Rita Population Adjustment>--Linked to County Attributes–Total U.S., 1969–2012 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch," released April 2014, based on the January 2014 submission. Available at [www.seer.cancer.gov]. [2]

Takahashi, K., Yokoyama, T., and Tango, T. (2015), "FlexScan User Guide for Version 3.0," [online]. Available at [http://www.niph.go.jp/soshiki/gijutsu/index_e.html]. [3]

Tango, T., and Takahaski, K. (2012), "A Flexible Scan Statistic with a Restricted Likelihood Ratio for Detecting Disease Clusters," *Statistics in Medicine*, 31, 4207–4218. doi:10.1002/sim.5478. [3]

U.S. Census Bureau (2011), Age and Sex Composition: 2010, 2010 Census Briefs. Available at [https://www.census.gov/prod/cen2010/briefs/c2010br-03.pdf]. [8]

——— (2016a), "Poverty Thresholds," [online]. Available at [https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-thresholds.html]. [2,7]

——— (2016b), "County Population by Characteristics Datasets: 2010–2016," [online]. Available at [https://www.census.gov/data/datasets/2016/demo/popest/counties-detailed-detail.html]. [2,7]

——— (2016c), "Educational Attainment in the United States: 2000 [and 2010]," [online]. Available at [https://www.census.gov/data/tables/2010/demo/educational-attainment/cps-detailed-tables.html]. [2,7]

——— (2016d), "Population and Housing Unit Estimates," [online]. Available at [https://www.census.gov/programs-surveys/popest.html]. [2,7]