Superfund Locations and Potential Associations with Cancer Incidence in Florida

Alexander Kirpich\textsuperscript{a} and Emily Leary\textsuperscript{b}

\textsuperscript{a}Department of Molecular Genetics & Microbiology, College of Medicine, University of Florida, Gainesville, FL; \textsuperscript{b}School of Medicine, University of Missouri, Columbia, MO

\textbf{ABSTRACT}
Uncontrolled hazardous wastes sites have the potential to adversely impact human health and damage or disrupt ecological systems and the greater environment. Four decades have passed since the Superfund law was enacted, allowing increased exposure time to these potential health hazards while also allowing advancement of analysis techniques. Florida has the sixth highest number of Superfund sites in the US and, in 2016, Florida was projected to have the second largest number of new cancer cases in the US. We explore statewide cancer incidence in Florida from 1986 to 2010 to determine if differences or associations exist in counties containing Superfund sites compared to counties that do not. To investigate potential environmental associations with cancer incidence, results using spatial and nonspatial mixed models were compared. Using a Poisson–Gamma mixture model, our results provide some evidence of an association between cancer incidence rates and Superfund site hazard levels, as well as proxy measures of water contamination around Superfund sites. In addition, results build upon previously observed gender differences in cancer incidence rates and further indicate spatial differences for cancer incidence. Heterogeneity among cancer incidence rates were observed across Florida with some mild association with Superfund exposure proxies.

\textbf{Introduction}
In the past 20 years, spatial analysis techniques have greatly advanced. The software capabilities for these techniques have greatly expanded as well. These new analytic tools allow more precise estimation methods and analyses that produce higher quality modeling results. In particular, these advanced techniques can be used to improve upon the existing body of literature that investigates associations between cancer incidence and environmental exposures. Most of the previous research investigating these relationships was conducted in the 1980s and early 1990s. Only a few recent studies have further explored these relationships, none of which have investigated nonpediatric cancer incidence. In addition, a longer time period of recorded cancer incidence and exposure data allow better characterization of exposures for the known long latency period prior to cancer incidence (Nadler 2014). Additional historical data and current advanced analysis techniques available to researchers provides an opportunity to restudy these relationships and consider trends over a longer time period.

When considering adverse environmental exposures, “Superfund” is probably one of the most recognized terms to describe the scale and extent of these exposure types. On December 11, 1980, U.S. President Jimmy Carter signed into law the Comprehensive Environmental Response, Compensation and Liability Act (CERCLA), known informally as Superfund. Once enacted, the Superfund law provided the U.S. Environmental Protection Agency (EPA) the legal authority to conduct clean-up efforts for uncontrolled hazardous waste sites and spills (EPA 2014). According to the EPA (2014), these sites include those “abandoned, accidentally spilled, and illegally dumped hazardous wastes that are determined to pose current or future threats to humans or the environment.” Long-term remediation efforts are also authorized by the Superfund law and conducted in areas where hazardous waste release occurred “through years of inadequate or illegal waste management” (EPA 2014).

There are multiple environmental factors to cancer risk and these include indoor and outdoor air pollution, soil contamination, and drinking water contamination (Boffetta and Nyberg 2003). In contrast to the large amount of work on air pollution and health risk, particularly lung cancer (Raaschou-Nielsen et al. 2013; Hamra et al. 2014), our focus for this study is on soil contamination and drinking water contamination exposures, common avenues of exposure for hazardous wastes, using Superfund sites as the environmental factor of interest. A comprehensive review of the literature on implications of hazardous
waste was published (Johnson and DeRosa 1997) and concluded that in spite of inherent study limitations, some studies have detected associations between cancer and exposure to materials similar to those that are found at hazardous waste sites. Older studies from the 1980s found increased frequency of gastrointestinal, esophageal, stomach, colon, lung, bladder, large intestine, and rectal cancers in counties containing hazardous waste sites (Najem et al. 1983; Budnick et al. 1984; Griffith et al. 1989). More current studies have continued to explore cancer incidence or other health impacts of hazardous waste exposure on the national or local level (Johnson 1995; Davies 2006; Florida Department of Health 2012a). Statewide cancer impacts and hazardous waste site exposure in Florida have previously been explored in one study, which focused on childhood cancers; results indicated some evidence of spatial clustering focused within 5–10 miles (8–16 km) of Superfund sites and not in areas in closer proximity to the Superfund site, although health data were aggregated to the geographic centroid of each census tract and may not reflect the real cluster location (Kearney 2008). A series of studies have been published that explore the pediatric cancer incidence in Florida and potential spatial clustering (Heaton 2014; Amin et al. 2014; Wang and Rodriguez 2014; Lawson and Rotejanaprasert 2014; Zhang, Lim, and Maiti 2014). For this series of studies, data were obtained from the Florida Association of Pediatric Tumor Program (FAPTR) and explored only pediatric onset cancers. Results indicate evidence of spatial clustering of pediatric cancer incidence; however, different areas were identified as potential clusters across studies. Further, in Heaton (2014), spatial differences in pediatric cancer incidence were observed by race but this was not observed in the other studies in this series.

Attributing pediatric cancers to environmental causes is difficult because pediatric cancer is rare but also because the causal linkage between environmental exposure pediatric cancer is unclear (National Cancer Institute 2016). However, environmental causes can be considered for adult cancers because the older age supports the longer latency needed. Nadler and Zurbenko (2014) provided a nice description of the approximate latency period from cancer initiation to diagnosis for 34 cancer types. Cancer latency periods range from 2.2 years for chronic lymphocytic leukemia to 56.8 years for ascending colon cancer. Unfortunately because location history is problematic to obtain, inference is still difficult.

Uncontrolled hazardous waste sites have the potential to adversely impact human health and damage or disrupt ecological systems and the greater environment. Specifically, improperly stored or damaged containers of hazardous wastes can contribute to overall toxic exposures through the food chain or through intakes as part of a water supply for human or agricultural consumption (Hendryx et al. 2012). The reasons for identifying potential Superfund sites are those designated by the Superfund law as (i) either causing or contributing to an increase in mortality or irreversible or incapacitating illness, (ii) posing a substantial present or potential hazard to human health, or (iii) when materials hazardous to humans or to the environment were improperly treated, stored, transported, disposed of, or otherwise managed (TOXMAP 2012a). More than 800 materials are designated as hazardous and nearly 100 additional materials are identified as potentially hazardous (TOXMAP 2012b, 2012c).

Four decades have passed since the Superfund law was enacted, which has allowed for more historical data to be collected regarding these potential health hazards. During this same time frame, analysis techniques—particularly geographic information system techniques (GIS)—have advanced and allow a more detailed investigation of these relationships. In this study, the association between all-cause cancer incidence in Florida and locations of Superfund sites are explored, using proxy measures of exposure to hazardous chemicals and various demographic variables as predictors.

**Methods**

**Data**

A publicly available information database for cancer incidence in Florida, called the STAT CD, was requested from the Florida Cancer Data System (FCDS) and available demographic covariates were age, sex, and county of residence (FCDS (Florida Cancer Data System) 2012). Because of the overwhelming white and non-Hispanic cancer incidence numbers in these data (>90% for both), these analyses did not separate out race and ethnicity demographics. Pediatric cancers were excluded from analysis because of the likely nonenvironmental cause; it is widely believed that children would not have the needed prolonged exposure to environmental toxins or chemicals to cause environmentally induced cancers (National Cancer Institute 2016). Age at the time of cancer incidence was categorized into groups, while, due to identifiability concerns, diagnosis year was grouped into 5-year intervals for the years between 1986 and 2010, inclusive. Due to privacy concerns and limited data availability, only county-level information was considered in this manuscript. University of Florida Health Science Center and University of Missouri Institution Review Board approval was obtained and nearly 2.4 million records were used for analysis (Tables 1 and 2).

We adjust the raw, total, all-cause cancer incidence counts to account for the total Florida population. There are two main methods of adjustments: the direct method and the indirect method (Naing 2000). The direct method of adjustment applies strata (here the 5-year intervals or age groups) observed in the population to a reference or standard population, here the 2000 US Census numbers for comparison to the Florida Cancer Data System rates (Florida Cancer Data System 2016). The direct method for adjustment results in the number of expected cancers in the reference population. The indirect method for adjustment is similar but observed rates are compared to the populations of interest, to result in the observed number of deaths in each population of interest. The direct method was used here.
to compare rates easily to results from the Florida Cancer Data System (Table 3).

We aggregated data for Florida Superfund sites and hazards (TOXMAP 2012a), Florida shape files for GIS techniques (U.S. Census Bureau 2008), socioeconomic status data (Florida CHARTS 2015), census data (U.S. Census Bureau 2008; Florida Department of Economic and Demographic Research 2016), and data from water reservoirs (ESRI 2012).

Before any hazardous waste site can be classified as a Superfund site, it must first be placed on the National Priorities List (NPL) in the U.S. Federal Register. To be placed on the NPL, it must first be placed on the National Priorities List for more details on OK. Ordinary kriging produces spatial predictions of the HRS and these predictions were located at each selected points. The reader is referred to supplementary materials for more details on OK. Ordinary kriging produces spatial predictions of the HRS and these predictions were located at each county’s centroid. For the spatial models, these spatial predictions were used as the HRS value for each county (Figure 1A). For nonspatial models, total county-level HRS was computed by summing each HRS score for all Superfund sites located in

| Table 2. Raw total, all-cause cancer incidence rates by age group and gender, 1986–2010. Total Rate refers to the rate per 100,000. |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| Age in years | <20 | 20–24 | 25–34 | 35–44 | 45–54 | 55–64 | 65–74 | 75–84 | 85+ | Total |
| Male | 8980 | 3932 | 17,216 | 38,832 | 104,354 | 238,303 | 423,224 | 320,631 | 84,416 | 1,239,888 |
| Female | 7992 | 6248 | 32,217 | 76,615 | 143,666 | 213,103 | 311,096 | 267,958 | 98,585 | 1,157,480 |
| Total | 16,972 | 10,180 | 49,433 | 115,447 | 248,020 | 451,406 | 734,320 | 586,583 | 183,001 | 2,397,368 |
| Total Rate | 0.0042 | 0.0110 | 0.0237 | 0.0465 | 0.0978 | 0.2895 | 0.5056 | 0.3747 | 0.5524 | 0.1500 |

was included, as a method to determine if there may be an association with cancer incidence and either previous exposure or continued exposure after remediation. We assume that possible confounding influences that contribute to cancer incidence are relatively homogenous across Florida.

Seventy-seven Superfund sites with a final, proposed, or deleted status are located in the state of Florida in 22 different counties (Figure 1A). The EPA defines a hazardous ranking score (HRS) using a numerical value from 0 to 100, with larger values indicating a larger potential health hazard to humans or the environment. The HRS is assigned by the EPA to each Superfund site, regardless of site status, and characterizes the combination of the potential hazard of the location, the possibility that a site has released or has the potential to release hazardous substances into the environment, characteristics of the waste (e.g., toxicity and waste quantity), and people or sensitive environments that could be affected by a release. Exposures to hazardous wastes for each Florida county were assigned using ordinary kriging of the HRS. Ordinary kriging (OK) is a method that is widely used for prediction of geostatistical data and assumes a constant mean in the neighborhood of each prediction. Values are predicted based on the known values at collected points. The reader is referred to supplementary materials for more details on OK. Ordinary kriging produces spatial predictions of the HRS and these predictions were located at each county’s centroid. For the spatial models, these spatial predictions were used as the HRS value for each county (Figure 1A). For nonspatial models, total county-level HRS was computed by summing each HRS score for all Superfund sites located in

| Table 3. Direct adjusted cancer incidence rates per 100,000 people, adjusted by age and gender, for each Florida county. |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| County | Cancer incidence rates | County | Cancer incidence rates | County | Cancer incidence rates |
| Alachua | 0.021522654 | Hardee | 0.0255996 | Okeechobee | 0.033823139 |
| Baker | 0.020349851 | Hendry | 0.02063085 | Orange | 0.021266464 |
| Bay | 0.025180772 | Hernando | 0.045675928 | Osceola | 0.020161342 |
| Bradford | 0.022905491 | Highlands | 0.04664726 | Palm Beach | 0.03725483 |
| Brevard | 0.03422357 | Hillsborough | 0.02374526 | Pasco | 0.03953705 |
| Broward | 0.022992687 | Holmes | 0.01657693 | Pinellas | 0.036756757 |
| Calhoun | 0.022345838 | Indian River | 0.04376571 | Polk | 0.030543017 |
| Charlotte | 0.044022034 | Jackson | 0.017604018 | Putnam | 0.032083605 |
| Citrus | 0.044124185 | Jefferson | 0.023405109 | Santa Rosa | 0.022175452 |
| Clay | 0.01981876 | Lafayette | 0.017269723 | Sarasota | 0.04513641 |
| Collier | 0.034963723 | Lake | 0.042589947 | Seminole | 0.0201223 |
| Columbia | 0.02480613 | Lee | 0.03567354 | St. Johns | 0.0263026 |
| Desoto | 0.022912272 | Leon | 0.017304542 | St. Lucie | 0.02596237 |
| Dixie | 0.030790716 | Levy | 0.032212975 | Sumter | 0.03514368 |
| Duval | 0.023550987 | Liberty | 0.022479976 | Suwannee | 0.028363052 |
| Escambia | 0.026277658 | Madison | 0.020280218 | Taylor | 0.02643739 |
| Flagler | 0.036337892 | Manatee | 0.034252666 | Union | 0.04230195 |
| Franklin | 0.030980032 | Marion | 0.039373031 | Volusia | 0.03637842 |
| Gadsden | 0.02810457 | Martin | 0.04284823 | Wakulla | 0.022475281 |
| Gilchrist | 0.027264755 | Miami-Dade | 0.023601532 | Walton | 0.019586739 |
| Glades | 0.019747373 | Monroe | 0.029481585 | Washington | 0.020003322 |
| Gulf | 0.028681152 | Nassau | 0.025367375 | | |
| Hamilton | 0.020036213 | Okaloosa | 0.02328763 | | |
Figure 1. (a) Locations of Superfund sites in Florida with the spatial prediction for county-level hazardous ranking score (HRS). (b) Locations of surface water reservoirs and National Weather Service (NWS) regions in Florida.

A county. Both of these measures can be considered a proxy for the level of hazard associated with Superfund site(s) within and around each county.

An important factor when considering hazardous waste and toxic materials is their ability to travel from the original waste site; one such method of travel is through water sources (EPA 1992). Because of the higher potential for hazardous wastes in Florida to travel through water—due to Florida’s swampy, wet environment—the proportion of surface water contained within a 1 km area around each Superfund site was calculated and rescaled. As shown in Figure 1B, most Superfund sites (black dots) in Florida have some type of surface water reservoir located nearby (blue areas). For Superfund sites near county lines, this proportion of surface water was considered part of the county in which the Superfund site was located. The aggregated rescaled proportion for all Superfund sites located within a county’s borders was used as the proxy for water exposure of hazardous wastes for the entire county.

Nonspatial Models

Cancer incidence counts were modeled using a Poisson–Gamma mixture model. The Poisson–Gamma mixture model is equivalent to the negative binomial generalized linear model and allows a more flexible model fit because it accounts for overdispersion. The negative binomial model is formulated as a mixture of the Poisson and gamma distributions and has the following form:

\[
Y \mid Z \sim \text{Poisson}(\mu Z)
\]

\[
Z \sim \text{Gamma}(\alpha, 1/\alpha)
\]

\[
= \frac{\alpha^\alpha}{\Gamma(\alpha)} e^{-\alpha} e^{-\alpha Z}
\]

where *Y* is the response of interest that can either be count data or rate data. Models with count data must include an offset term with total population counts; this ensures that the counts are scaled to the total counts. To ensure the required flexibility for comparison to the standard Poisson model, *Z* is assumed to have a gamma distribution as previously defined.

As defined, the marginal distribution of *Y* has the probability mass function:

\[
p(y | \mu, \alpha) = \frac{\Gamma(y + \alpha)}{y! \Gamma(\alpha)} \left( \frac{\mu}{\mu + \alpha} \right)^y \left( \frac{\alpha}{\mu + \alpha} \right)^\alpha, \quad y = 0, 1, 2, 3 \ldots
\]

For integer values \(\alpha = m\), the mass function (1) simplifies to the negative binomial mass function:

\[
p(y | \mu, m) = \binom{m + y - 1}{m - 1} (1 - p)^y p^m, \quad y = 0, 1, 2, 3 \ldots
\]

For fixed \(m\) (or \(\alpha\)), the model follows the standard generalized linear equation framework with the link function, \(g(E[Y])\). The linear predictor has the form \(\eta = g(E[Y]) = X^T \beta\), where \(X^T\) is the transposed vector of predictors and \(\beta\) is the vector of model coefficients. For arbitrary \(\alpha\), the mean and variance have the form:

\[
E[Y] = \mu \text{ and } \text{var}[Y] = \mu \left(1 + \frac{\mu}{\alpha} \right)
\]

This model was fit (see the supplementary materials) using a population offset from the 2000 U.S. Census (U.S. Census Bureau), and thus all interpretations will be in terms of cancer incidence rates. Using the 2000 U.S. Census numbers, rather than the 2010 U.S. Census numbers, was intended to provide uniform comparison to the Florida Cancer Data System rates, which use the 2000 U.S. Census numbers (Florida Cancer Data System 2016). For the nonspatial analysis, the 67 counties within Florida could not be used as a spatial proxy due to the large number of counties, which when included as an indicator variable, produced model instability due to the large number of parameters. However, the National Weather Service (NWS) defines seven different weather regions within Florida, which are defined using county boundaries within Florida (Figure 1b).
These NWS regions were used as spatial proxies in all nonspatial analyses and are composed of the Jacksonville region (JAX), Keys region (KEY), Miami region (MFL), Melbourne region (MLB), Mobile region (MOB), Tallahassee region (TAE), and the Tampa Bay region (TBW). The summary information for Superfund site locations by NWS region is provided in Table 4.

Categorical covariates included in the nonspatial model were age group, gender, diagnosis period, presence of Superfund indicator, and NWS region. Continuous predictor and covariates were median income, the number of proposed and final Superfund sites, total county-level HRS, and county-level water exposure. Median income was used to control the potential confounding from socioeconomic status. Demographic covariates, such as gender and age group, were included to account for the known differences in cancer incidence by gender and age (e.g., cancer incidence increases for older age groups). Diagnosis period was also included in the analysis as it can be considered a measure of latency or proxy for medical care during this large time frame.

**Spatial Models**

A spatial simultaneous autoregressive error model (SAR Error Model) was used for spatial modeling. These model equations had the form:

\[ Y = X\beta + u \text{ where } u = \lambda W u + \epsilon. \]

Or equivalently,

\[(I - \lambda W) u = \epsilon \text{ where } E[\epsilon] = 0 \text{ and } \text{var}[\epsilon] = \sigma^2 I.\]

Here, \( W \) was the matrix of the boundary weights, \( Y \) was the response vector, \( X \) is the matrix of predictors, \( I \) is the identity matrix, and \( \epsilon \) is the vector of normally distributed errors. The parameters \( \lambda \) and \( \beta \) were estimated (see supplementary materials). The county-level spatial structure for Florida was described using a boundary matrix \( W \) of dimension 67 by 67 representing the number of Florida counties—using two different types of spatial weights to represent the spatial relationships between counties: (1) binary 0-1 queen weights and (2) standardized queen weights. Binary 0-1 queen weights for a given county are stored in a single row of the weights matrix. The weight value is equal to 1 if counties have at least one shared boundary point and 0 otherwise. Standardized queen weights are obtained from binary 0-1 queen weights by dividing each element in the row by the sum of all elements in that row. The individual elements of \( W \) are denoted by \( w_{ij} \) for \( i = 1, 2, \ldots, N \) and \( j = 1, 2, \ldots, N \) where \( N = 67 \) corresponds to the number of counties in Florida. Since spatial models are constructed so that only a single rate can be used for each spatial element (here, county), age and diagnosis year were included in these spatial models by adjusting the cancer incidence rates for each county by both age and diagnosis year, using US Census Data for Florida from 2000 as previously discussed. County-level exposure variables included in the models included the ordinary kriged HRS, county-level water exposure, the number of proposed and final Superfund sites for the given county and the number of deleted Superfund sites.

To investigate spatial heterogeneity, the Global Moran’s I statistic was calculated for the spatial models to determine the presence of any spatial autocorrelation (spatial relationship) between the cancer incidence rate and county. Global Moran’s I is sensitive to global spatial autocorrelation and has the form:

\[ I = \frac{N}{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} (X_i - \bar{X}) (X_j - \bar{X})}{\sum_{i=1}^{N} (X_i - \bar{X})^2}. \]

Here, \( X_i \) are the values for the spatial objects and \( w_{ij} \) are the weights from the \( 67 \times 67 \) weight matrix, \( W \), for \( i = 1, 2, \ldots, 67 \) and \( j = 1, 2, \ldots, 67 \). The value \( \bar{X} \) corresponds to the average value of \( X_i \) across all spatial elements, that is, \( \bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i \). The reader is referred to supplementary materials for more details on the Global Moran’s I statistic.

Although similar to the Global Moran’s I, Geary’s C statistic was also used to focus on relationships at the local level. The statistic has the form

\[ C = \frac{(N - 1) \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} (X_i - X_j)^2}{2 \left( \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} \right) \sum_{i=1}^{N} (X_i - \bar{X})^2}. \]

The reader is referred to supplementary materials for more details on Geary’s C statistic.

To further investigate the observed spatial relationships, local Getis-Ord general G and local Moran’s I tests were conducted. Local Getis-Ord \( G_i \) are used to identify clusters with high and low values. Local Getis-Ord \( G_i \) statistics for each spatial object \( i \) have the form:

\[ G_i = \frac{\sum_{j=1}^{N} w_{ij} X_j - \bar{X} \sum_{j=1}^{N} w_{ij}}{\sqrt{\frac{\sum_{i=1}^{N} X_i^2}{N} - \bar{X}^2} \sqrt{\frac{\sum_{j=1}^{N} w_{ij}^2 - \left( \sum_{j=1}^{N} w_{ij} \right)^2}{N-1}}}. \]

Anselin Local Moran’s I statistics were introduced to detect local spatial autocorrelation. The formula for each spatial object \( i \) has the form:

\[ I_i = \frac{X_i - \bar{X}}{\left[ \frac{1}{N-1} \sum_{j=1, j \neq i}^{N} (X_j - \bar{X})^2 \right] \sum_{j=1, j \neq i}^{N} w_{ij} (X_j - \bar{X})}. \]

The local Getis-Ord general G tests for local clustering of cancer incidence rates will indicate any difference from expected cancer incidence rates. In addition, the Local Moran’s I statistic helps to identify if a county has similarly high or low cancer incidence rates with the surrounding counties and can be helpful to indicate potential clusters of either high or low cancer incidence rates. Used together with Getis-Ord general G, results can identify potential clusters and cancer incidence rates that could be considered “outliers” compared to rates in surrounding counties. The reader is referred to the supplementary materials for more details on Local Getis-Ord G and Anselin Local Moran’s I, statistics.

**Results**

Results from both the spatial and nonspatial models indicated that spatial differences are observed in adult cancer incidence in Florida from 1986–2010. Further, nonspatial models indicated that increased county-level HRS, county-level water exposure for hazardous wastes and being male were associated with increased rates for adult cancer incidence.
Frequency of Superfund status by NWS regions

| Site status     | JAX  | KEY  | MFL  | MLB  | MOB  | TAE  | TBW  |
|-----------------|------|------|------|------|------|------|------|
| Proposed        | 1    | 0    | 0    | 1    | 0    | 0    | 2    |
| Final           | 7    | 0    | 12   | 9    | 5    | 4    | 13   |
| Deleted         | 2    | 0    | 11   | 0    | 3    | 6    |
| All             | 10   | 0    | 23   | 10   | 8    | 5    | 21   |

Combined HRS by Superfund status and NWS regions

| Site status     | JAX  | KEY  | MFL  | MLB  | MOB  | TAE  | TBW  |
|-----------------|------|------|------|------|------|------|------|
| Proposed        | 70.71| 0    | 0    | 50   | 0    | 0    | 99.98|
| Final           | 274.4| 0    | 503.09| 380.39| 245.79| 169.36| 626.73|
| Deleted         | 75.77| 0    | 423.58| 0    | 124.32| 37.61| 185.05|
| All             | 420.88| 0    | 926.67| 430.39| 370.11| 206.97| 911.76|

Table 5. The results of the Poisson-Gamma mixture model fit.

| Coefficient name | Estimate | Std. error | 95% confidence intervals | p-value | Sig. |
|------------------|----------|------------|--------------------------|---------|------|
| Intercept        | −6.4584  | 0.0409     | −6.5404                  | < 0.001 | ***  |
| Hazardous Ranking Score | 0.0017 | 0.0006 | 0.0004 | 0.0029 | 0.0091 | **  |
| Water 1km Buffer | 0.0002  | 0.0001     | 0                        | 0.0004  | 0.0284 | *   |
| Proposed and Final | −0.0961 | 0.0318   | −0.1587                 | < 0.0334| 0.0225 | **  |
| Deleted          | −0.0313 | 0.0216     | −0.0737                 | 0.0112  | 0.1478 | .    |
| Gender (Female)  | −0.0197 | 0.0101     | −0.0398                 | 0.0005  | 0.0508 | .    |
| 25–34 years old | 0.7865  | 0.0255     | 0.7361                  | 0.837   | < 0.055| ***  |
| 35–44 years old | 1.5359  | 0.0248     | 1.505                   | 1.6208  | < 0.001| ***  |
| 45–54 years old | 2.4549  | 0.0244     | 2.4068                  | 2.503   | < 0.001| ***  |
| 55–64 years old | 3.2561  | 0.0242     | 3.2083                  | 3.3039  | < 0.001| ***  |
| 65–74 years old | 3.8409  | 0.0241     | 3.7933                  | 3.8885  | < 0.001| ***  |
| 75–84 years old | 4.0223  | 0.0242     | 3.9746                  | 4.0701  | < 0.001| ***  |
| ≥85 years old   | 3.9899  | 0.0247     | 3.941                   | 4.0387  | < 0.001| ***  |
| Indicator of Superfund Site | 0.0634 | 0.0151 | 0.0339 | 0.0929 | < 0.001| ***  |
| Diagnosis 1991–1995 | 0.1497 | 0.0161 | 0.118  | 0.1815 | < 0.001| ***  |
| Diagnosis 1996–2000 | 0.0412 | 0.0161 | 0.0096 | 0.0728 | 0.0106 | *    |
| Diagnosis 2001–2005 | 0.0234 | 0.016 | −0.0081 | 0.055  | 0.1443 | .    |
| Diagnosis 2006–2010 | −0.0158 | 0.016 | −0.0474 | 0.0158 | 0.323  | .    |
| JAX Region      | 0.2335  | 0.0167     | 0.1808                  | 0.2462  | < 0.001| ***  |
| KEY Region      | 0.1839  | 0.0434     | 0.0993                  | 0.2699  | < 0.001| ***  |
| MFL Region      | 0.0465  | 0.0247     | −0.0022                 | 0.0953  | 0.0598 | .    |
| MLB Region      | 0.1779  | 0.0197     | 0.1394                  | 0.2165  | < 0.001| ***  |
| MOB Region      | 0.0656  | 0.0283     | 0.0107                  | 0.1209  | 0.0203 | *    |
| TBW Region      | 0.0704  | 0.0162     | 0.0137                  | 0.2021  | < 0.001| ***  |
| Median Income   | 0.0019  | 0.001      | −0.0001                 | 0.0039  | 0.0599 | .    |

Nonspatial Models

The nonspatial model fit results are provided in Table 5. In the nonspatial Poisson–Gamma mixture model using NWS regions as spatial proxies, each age group category was observed to be significantly associated with cancer incidence ($p < 0.010$). This significance reflects the higher cancer incidence observed as a population ages. Being female was inversely associated with cancer incidence, with $p$ value equal to 0.0508, indicating a potential protective effect against cancer incidence for females, or reduced exposure. Of the exposure variables, county-level water exposure for hazardous wastes and total county-level HRS were positively associated with cancer incidence ($p = 0.028$ and $p = 0.009$, respectively). The number of proposed and final Superfund sites were also significantly associated with cancer incidence ($p = 0.0025$), indicating that counties with a greater number of proposed and final sites were related to increasing cancer incidence. However, the direct opposite relationship that counties with fewer Superfund sites—of all types—also related to increased cancer incidence. In Kearney (2008), most Superfund sites near a cancer cluster were located within 5–10 miles rather than within 0–1 mile; perhaps that study, as well as this study, are observing effects of the spatial misalignment problem—data are not measured at the same spatial level and the misalignment propagates error throughout the analysis.

In addition, evidence of extreme spatial heterogeneity among cancer incidence rates was observed between the seven NWS geographic regions. Because of the larger size and few Superfund sites, the Tallahassee region (TAE) was used as the reference region for all nonspatial analyses. Of the remaining six regions, all other regions demonstrated strong positive association with increased cancer incidence rates in comparison to Tallahassee region (TAE) region ($p < 0.05$), with the exception of the Miami region (MFL). The strongest associations were observed in the Jacksonville region (JAX), the Melbourne region (MLB), and the Tampa Bay region (TBW). These relationships likely reflect that these regions contain the most Superfund sites. The covariate for median income was on the boundary of significance ($p = 0.060$).
**Spatial Models**

Spatial models were constructed separately for each gender and similar results were obtained using each type of spatial weight. However, no evidence of a significant association between environmental exposures and cancer incidence rates was observed.

Because of the observed regional differences using the non-spatial models, the Global Moran’s I and Geary’s C statistics were calculated for the spatial data, to determine the presence of any spatial autocorrelation between the cancer incidence rate and county. Both Moran’s I and Geary’s C presented evidence of spatial autocorrelation with p values ranging from $p < 0.0001$ (Moran’s I, males, standardized queen weights) to $p = 0.07$ (Geary’s C, females, binary queen weights). The results depended on the test, gender, and weights used. The tests for global spatial autocorrelation were significant for both males and females; smaller p values were observed for males with the largest value, $p = 0.003$.

To investigate further, the local Getis-Ord general G and local Moran’s I tests were conducted. Both tests produced similar patterns for males and females, with the exception of a few counties (Figure 2). Similar to the conclusions from Global Moran’s I and Geary’s C, these results indicate that cancer incidence rates for males have stronger spatial heterogeneities. That provides evidence that cancer rates are indeed different between counties and identifies the patterns in the data. Considering the Local Moran’s I test together with the Getis-Ord general G, overall higher geographic variability was observed for male rates indicating spatial differences in cancer incidence rates by gender (Figure 2).

**Discussion**

Using nonspatial analysis techniques, our results suggest a potential association between increased HRS score and increased proxy water exposure of hazardous wastes around Superfund sites with increased adult cancer incidence rates. Further, presence of a Superfund site in a county was associated with an increase in cancer incidence rates. Relationships between gender and age covariates, specifically being older and male, are associated with increased cancer incidence. Results from this research that show associations with cancer incidence rates between age and the male sex are not unexpected based on the typical longer latency period needed for cancer onset (Nadler 2014) and established gender differences in cancer incidences (Cook et al. 2009; Edgern et al. 2012). These results are important as almost 37% of the total population in Florida is 50 years of age or older and the resident median age for Florida is 41.2 years old, higher than the national median age of 37.4 years old (U.S. Census Bureau 2015). Unexpectedly, higher median income was associated with a higher cancer incidence rate but this is perhaps an indication that diagnostic procedures might be more frequent or of a higher quality for people or families with higher income. This result may reflect that the NWS regions that contain the largest numbers of Superfund sites are also those regions with both wealthy and urban populations.

In addition, heterogeneity of cancer incidence was identified in all NWS regions without exception and NWS regions with the largest numbers of Superfund sites (JAX, MLB, TBW) had the strongest associations. Confirming results from the nonspatial model, Global Moran’s I provided evidence of spatial autocorrelation for both types of spatial weights and sex. Similar results were observed using Geary’s C statistic indicating that similar cancer incidence rates cluster spatially. The Getis-Ord G and Local Moran’s I results indicated that there is stronger spatial heterogeneity among males, although one high-high cluster was observed in one county for both males and females.

The spatial models did not find evidence of a significant association. This result could be due to the structure of the spatial model, as only one response was allowed for each spatial unit (e.g. county). The direct rate adjustment could have diluted a signal in the data, the signal that was observed in the non-spatial model. One suggestion is that future studies consider techniques to better ameliorate spatial misalignment in the data, such as using block kriging for county level estimates.

Although previous research has indicated an association between gender and increased cancer incidence (Cook et al. 2009), results from this analysis indicate that in addition, there are spatial differences between genders for cancer incidence. These results build on the spatial differences observed in Goli et al. (2013) by identifying these spatial differences for a Western population while also investigating their associations with environmental exposures related to Superfund site locations. Spatial heterogeneity is important for future epidemiological work into identifying environmental causes of cancer clusters as well as to inform environmental regulations and policy.

A perplexing result from the nonspatial analysis is the observed relationship between decreasing numbers of proposed and deleted Superfund sites and increased cancer incidence, while a protective relationship is observed when considering all types of Superfund sites (proposed, deleted, and final) and cancer incidence. In addition, counties with small numbers of Superfund sites were found to have high cancer incidence rates. This could be due to location or material characteristics of the final Superfund sites. For example, Miami-Dade county has experienced massive immigration from inside and outside the US, in the past few decades. This has caused the population in this area to dramatically increase. Because of this population inflow, environmental exposures to toxic chemicals in this area is less likely for the new residents. In addition, the influx of new residents increased total population counts, possibly confounding cancer rates for this area. During the same time frame, immigration to the other parts of Florida (especially North Florida) was more moderate.

As in most research studies, limitations of the data and analysis are problematic. The demographics in the cancer incidence database for Florida contains an overwhelming number of white, non-Hispanic persons (>90%). These values are different from the overall population of Florida which is more diverse with 55% white, non-Hispanic and nearly 25% white, Hispanic (US Census Bureau 2015). Another limitation of the analysis is the privacy concern motivating aggregation to the county level. Although beyond the scope of this analysis, to investigate overall all-cause cancer incidence, further investigation into different categories of cancer incidence may reveal more about possible explanations for these observed spatial differences. For
example, differences in cancer incidence could be attributed to differences in occupation, residential location, risk behaviors such as tobacco use, or lifestyle differences such as staying indoors. Although further analysis and more detailed data, particularly exposure histories, are needed to better attribute what may be driving these spatial differences, the range of possible factors makes this a daunting task. The recent advances in spatial analysis techniques and software will become useful to further investigate cancer incidence and environmental hazardous waste exposure, when exposure history data are recorded and made available for study.

**Conclusions**

Results indicated potential association with environmental exposures related to Superfund sites and cancer incidence rates for Florida. In addition, evidence of spatial differences by gender for county-level cancer incidence rates in Florida were observed. More research is recommended on this topic to further identify and define these differences. Data availability creates the biggest challenge to further the science. Within the United States, cancer registries are defined and run by each state and they do not routinely include geographical or exposure histories that indicate length of residence at the time of cancer diagnosis, a thorough residential history, or an extremely detailed smoking and drinking exposure history. However, these exposure histories are essential to thoroughly investigate the effect of environmental exposures on cancer.

**Abbreviations**

- ACS American Cancer Society
- ArcGIS Geographic information system software produced by Environmental Systems Research Institute
- CERCLA Comprehensive Environmental Response, Compensation and Liability Act
- EPA U.S. Environmental Protection Agency
- FAPTP The Florida Association of Pediatric Tumor Program
- FCDS Florida Cancer Data System
- FDOH Florida Department of Health
- GIS Geographic Information System
- GLM Generalized Linear Model
- HRS Hazardous Ranking Score
- NPL National Priorities List
- NRC National Research Council
- NWS National Weather Service
- SAR Model Spatial Autoregressive Model

**Acknowledgments**

The Florida cancer incidence data used in this report were collected by the Florida Cancer Data system under contract with the Florida Department of Health. The views expressed herein are solely those of the author(s), and do not necessarily reflect those of the contractor of the Florida Department of Health.

The authors thank Dr. Xiaohui Xu and Dr. Jonathan D. Sugimoto for the project guidance and critical review of the analysis performed. In addition, the authors thank Dr. Wendell P. Cropper, Jr. for general guidance and early, critical review of the manuscript. The authors contributed equally to this article.

**Funding**

This research project was not funded by grant or any other source.

**Conflict of Interest Statement**

The authors state that they have no competing financial interests or conflict of interests in conducting this research.

**Supplementary Materials**

The supplementary materials contain detailed background information on the statistics and models used as well as a reference map of the 67 counties in Florida.
