Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____	_____
Sabah Munir	Date

The Application of Spatial Regression Analysis in Examining the Relationship Between Drinking Water Contaminants and Cancer Incidence Rates in Iowa and Illinois

By

Sabah Munir

Master of Science in Public Health (MSPH)

Department of Biostatistics & Bioinformatics

_____

Howard H. Chang, Ph.D.

Thesis Advisor

_____

Lance Waller, Ph.D.

Reader

The Application of Spatial Regression Analysis in Examining the Relationship Between Drinking Water Contaminants and Cancer Incidence Rates in Iowa and Illinois

By

Sabah Munir

Bachelor of Arts (BA)

Wartburg College

2019

Thesis Advisor: Howard H. Chang, Ph.D.

Reader: Lance Waller, Ph.D.

An abstract of

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science in Public Health

in Biostatistics & Bioinformatics

2021

Abstract

The Application of Spatial Regression Analysis in Examining the Relationship Between Drinking Water Contaminants and Cancer Incidence Rates in Iowa and Illinois
By Sabah Munir

Background: Past studies have examined the relationships between drinking water contaminants and cancer risk or incidence. However, many of these previous designs limited their scope to just one state, and the use of GIS techniques was restricted to only studying cancer incidence patterns. The current study sought to use spatial regression analysis to study the relationships between drinking water contaminants and cancer incidence in two Upper Midwest states (i.e., Iowa and Illinois).

Materials and Methods: Gaussian GAMs were used to conduct exploratory analyses between contaminants and cancer incidence rates while controlling for state and county urbanicity identifiers. SAR models helped study the associations between contaminants and cancer incidence rates while controlling for state and county urbanicity and examining the significance of spatial dependency on contaminant data.

Results: Both Illinois and Iowa reported fewer variations in concentration values for nitrates and arsenic. Higher radium values clustered in eastern Iowa and northern Illinois, and high haloacetic acids and total trihalomethane concentrations occurred in the southern halves of each state. Cancer incidence rates for all four cancers were varied across both states, and a cluster of higher prostate incidence was identified in western Iowa. Based on spatial analysis, for the incidence rates of both colon and rectal cancer and lung and bronchial cancer, radium yielded negative associations while total trihalomethanes yielded positive associations. Exploratory analysis also yielded significant positive associations between radium and female breast cancers and between haloacetic acids and lung and bronchial cancers. No significant associations were found between any contaminants and prostate cancer.

Discussion and Conclusions: The findings regarding radium's association with cancer incidence supported only past literature for female breast cancer. Likewise, the findings regarding haloacetic acids and total trihalomethanes partly supported past literature for colon and rectal and lung and bronchial cancers. Spatial regression was useful in identifying relationships between contaminant exposure and cancer incidence while controlling for state and urbanicity classifications, although only one model demonstrated significant second-order spatial dependence. The existing model can be expanded further by including additional variables, stratifying the dependent variable by socio-demographic factors, or including more states in the sample data.

The Application of Spatial Regression Analysis in Examining the Relationship Between Drinking Water Contaminants and Cancer Incidence Rates in Iowa and Illinois

By

Sabah Munir

Bachelor of Arts (BA)

Wartburg College

2019

Thesis Advisor: Howard H. Chang, Ph.D.

Thesis Reader: Lance Waller, Ph.D.

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science in Public Health

in Biostatistics & Bioinformatics

2021

Acknowledgements

I am deeply grateful to my thesis advisor, Dr. Howard Chang, for his mentorship, guidance, and support throughout this thesis process. His confidence in my ability to secure the data and ensure the feasibility of this thesis topic means a lot to me, and he provided me with a platform to engage in a process of self-discovery as a researcher. He challenged me to think in a more insightful manner about the research and to focus more on the larger picture of the findings. For this, I am deeply appreciative.

I also thank Dr. Lance Waller for reading my thesis manuscript and providing additional feedback on the draft. His suggestions enabled me to think more critically about how my own project relates and compares to past literature.

I am grateful to the Iowa Department of Natural Resources and Illinois Environmental Protection Agencies for their assistance in helping me identify and obtain the water data sources I needed to conduct this project.

I am also grateful to the faculty and staff of the BIOS department, from whom I have learned a lot more about the applications and theories behind biostatistics. Retrospectively, I also thank the faculty at Waverly-Shell Rock High School and Wartburg College for providing me with the opportunities to learn more about the biology and epidemiology of cancer, especially in Iowa.

Lastly, I thank my family for their love, support, and encouragement through my journey into biostatistics. They joined me as I discovered what the field of biostatistics is all about and supported me as I made the decision to study public health.

Table of Contents

1. Introduction

In the United States, incidence rates for most types of cancer have been decreasing or have remained stable since the year 1975 (Siegel et al., 2021). However, specific directions of change in cancer incidence can differ based on cancer site/type and socio-demographic factors. For example, male prostate and female breast cancer incidence rates grew in the 1980's and 1990's due to increased awareness and screening. Likewise, different patterns can be observed within a cancer type based on sex; between 1975 and 2017, female lung and bronchus cancer incidence rates increased as the corresponding rates for males decreased. This was partly due to changes in smoking prevalence, where the overall rate of smoking decreased but that some female birth cohorts reported increased smoking prevalence compared to preceding cohorts (Siegel et al., 2021).

The risk for cancer can also vary based on whether an individual lives in a rural area. Farmers were found to be at a lower risk for colon and lung cancers but at a higher risk for prostate cancer and for relatively less-common cancers (e.g., brain cancer) (Blair & Freeman, 2009). As found in the context of late-stage colorectal cancer, however, higher rates of late-stage cancer were less prevalent in urban locations and more associated with increased rates of non-local diagnoses. Individuals from rural locations were more likely to travel longer distances to access facilities equipped for cancer diagnoses, most of which are located in urban areas (Rushton et al., 2004). Therefore, urbanicity status does play some role in the incidence and severity of cancers.

Cancer development can be impacted by many exposure-related factors, one of which includes the widespread use of chlorine as a drinking water disinfectant (Evans et al., 2020). Despite this process reducing the risk of infection from water-borne pathogens, water chlorination also produces several byproduct compounds that are thought to be associated with negative health outcomes such as birth defects, miscarriages, and cancer (Evans et al., 2020). These chlorination

byproducts can be categorized into different groups, of which trihalomethanes and haloacetic acids comprise a majority of all chlorination byproducts (Gopal et al., 2007). Total trihalomethanes, or TTHMs, result from the reactions between hypochlorous acid (a compound derived from reactions between water and chlorine gas) and other organic compounds in water. Among the class of trihalomethanes, chloroform, bromodichloromethane, and bromoform are of specific interest with respect to their effects on cancer development. Bromoform in particular is associated with a small incremental lifetime risk for colon and rectal cancer (Dobaradaran et al., 2020), and exposure to high levels of chloroform is associated with an increased risk of breast cancer (Font-Ribera et al., 2018). Likewise, haloacetic acids, abbreviated as HAA or HAA5, come from reactions between chlorine and other organic compounds; dichloroacetic acid is of most concern in this category as a potential carcinogen. Past meta-analyses have indicated that the byproducts of water chlorination do not significantly change the relative risks for breast, colon, or lung cancers, but they do significantly increase the risk for rectal cancer (Morris, 1995).

Other potentially-carcinogenic chemicals can seep into drinking water sources from the soil; this includes radium, arsenic, and nitrates. Radium is often found in aquifer systems across the United States; it can seep from the soil into water ahead of its use for human consumption (Szabo et al., 2012). Levels of radium-226 above 5 pCi/L were associated with higher risks for lung cancer among males and breast cancer among females (Bean et al., 1982). Arsenic levels in groundwater samples, especially in the Upper Midwest, tend to be elevated due to the presence of glacial sediments in the soil (Erickson & Barnes, 2005). Higher exposure levels of arsenic were associated with higher risk ratios for prostate cancer, including aggressive cancers; it is also key to note that even at these higher levels, the arsenic concentration is still considered as low-level (Roh et al., 2017). Nitrate can naturally occur in soils as a result of bacteria, dying vegetation, or animal waste products, but it can

also come from fertilizers or as airborne industrial or automobile outputs (Manassaram et al., 2006). Because of its prevalence in soil, nitrate can contaminate groundwater, which serves as a major source of drinking water (Manassaram et al., 2006). Even at acceptable levels of nitrate, increased concentrations in drinking water are associated with an increased risk of colorectal cancer (Schullehner et al., 2018). Other studies have reported mixed findings. For example, no association was found between increasing concentrations of nitrates and the relative risks for different cancers (colon, breast, lung) among older Iowan women, but these nitrate levels showed a significant negative relationship with the risk for rectal cancers, suggesting a protective effect (Weyer et al., 2001).

In this project, we used a spatial simultaneous autoregressive model to examine cancer incidence rates in two states in the Upper Midwest: Iowa and Illinois. These two states were selected for characteristics that may be relevant to cancer risk. Both states were affected by geographic factors such as the Late Wisconsinan glacial drift and the Mid-Continent and Ozark Plateau Cambro-Ordovician aquifer system, which were implicated for higher groundwater arsenic and radium levels in the region (Erickson & Barnes, 2005; Szabo et al., 2012). In addition, these neighboring states are unique in that they are home to major urban areas while also being agriculture-heavy, and their shared border is defined by a major waterway (the Mississippi River) that is often a focus of health and environmental research (Jones et al., 2018).

The present analysis on these two states would allow us to examine the relationship between cancer incidence rates and the concentrations of certain drinking water contaminants while considering the role of spatial dependence. While past studies have also examined this relationship, many of them such as Blair and Freeman (2009) and Rushton et al. (2004) have focused on a single state. In addition, studies such as Mandal et al. (2009) and Rushton et al. (2004) also utilized GIS

techniques, but their analyses were limited to patterns in cancer incidence across one or more states. Our current analysis covers two states in examining both cancer incidence and variations in contaminant concentrations.

2. Materials and Methods

2.1 Data Sources

2.1.1 Drinking Water Contaminant Data

We obtained drinking water data between January 1990 and December 2017 that had been collected as part of the Environmental Protection Agency's Safe Drinking Water Information System (SDWIS). As required by the Safe Drinking Water Act, local agencies in each state regularly measure drinking water sources at various locations in state counties for contaminants and organisms, the results of which are stored in a federal database. Direct requests for water sample measurements were made to Iowa and Illinois authorities. Iowa water data from all 99 counties were obtained via email communication with a water supply environmental specialist at the Iowa Department of Natural Resources. Illinois water data from all 102 counties were obtained through a combination of an official FOIA request and email communication with a public service administrator at the Illinois EPA. The raw SDWIS datasets for Iowa contained 33,608 nitrate observations, 4,357 radium observations, 4,771 arsenic observations, 19,456 haloacetic acid observations, and 29,578 total trihalomethane observations. The raw SDWIS datasets for Illinois contained 49,070 nitrate observations, 14,250 radium observations, 20,912 arsenic observations, 70,531 haloacetic acid observations, and 74,708 total trihalomethane observations.

Within each contaminant type, all observations belonging to a given county were used to calculate a mean concentration value across the full 1990-2017 exposure window for that county. In addition, this method was used to calculate decade-specific mean concentration values, where the

year of collection was extracted as a new variable and used to flag whether the observation was recorded in the 1990's, the 2000's, or the 2010's prior to finding county-level means in each decade subset. Descriptive statistics of the county-level drinking water data are presented in Table 1, including the number of counties with available data and means, standard deviations, and ranges across counties.

Table 1. Descriptive statistics for county-level drinking water contaminant concentrations by state and exposure window

		Iowa			Illinois	
Contaminant By Exposure Window, Units	Sample Size	Mean (SD)	Range	Sample Size	Mean (SD)	Range
Nitrates, mg/L						
-All Years	99	2.21 (1.70)	0.07-6.87	99	1.56 (1.80)	0.01-8.15
-1990-1999	99	2.51 (2.02)	0.03-8.14	99	1.51 (1.69)	0-7.31
-2000-2009	93	2.00 (1.72)	0.02-6.87	99	1.41 (1.66)	0-7.61
-2010-2017	92	2.09 (1.71)	0.08-6.07	99	1.58 (1.89)	0-9.39
Radium, pCi/L						
-All Years	94	1.77 (1.16)	0.18-5.20	99	1.98 (1.52)	0.24-5.85
-1990-1999	90	1.66 (1.41)	0-6.10	27	6.10 (2.73)	1.50-13.00
-2000-2009	93	1.83 (1.16)	0.33-5.62	95	2.14 (1.85)	0-7.31
-2010-2017	92	1.38 (1.18)	0-4.10	99	1.71 (1.13)	0-5.22
Arsenic, μg/L						
-All Years	99	0.92 (1.29)	0-7.59	99	2.74 (4.08)	0-30.50
-1990-1999	99	0.24 (0.58)	0-3.23	98	2.22 (3.17)	0-14.80
-2000-2009	92	1.40 (1.66)	0-9.86	99	2.97 (5.78)	0-48.90
-2010-2017	91	1.31 (1.48)	0-7.84	99	2.48 (3.11)	0-14.50
Haloacetic Acids (HAA5), μg/L						
-All Years	99	11.40 (11.20)	0.45-54.10	99	16.10 (10.30)	0.72-43.30
-1990-1999	11	15.50 (7.77)	3.50-25.70	63	20.10 (19.80)	0-95.70
-2000-2009	99	11.10 (13.20)	0.00-66.90	99	17.40 (14.30)	0.61-93.90
-2010-2017	99	11.20 (10.00)	0.60-44.40	99	14.30 (8.49)	0.79-33.90
Total Trihalomethanes (TTHM), μg/L						
-All Years	99	23.10 (15.80)	1.50-79.20	102	37.30 (46.00)	3.03-425.00
-1990-1999	74	9.87 (15.80)	0-99.40	78	38.40 (25.50)	0-113.00
-2000-2009	99	26.00 (20.20)	0.52-84.60	102	44.70 (105.00)	1.95-1030.00
-2010-2017	99	25.70 (15.60)	1.66-64.20	102	29.70 (14.70)	3.62-62.20

2.1.2 Cancer Incidence Data

County-level cancer incidence rates were downloaded directly from the State Cancer Profiles public database of the National Institutes of Health (National Cancer Institute). Rates were provided as the latest five-year age-adjusted incidence rates per 100,000 based on data from 2013-2017.

County-specific rates were obtained for Iowa and Illinois based on specific cancer site (colon and rectum, lung and bronchus, breast [female only], or prostate), sex (both sexes, males, or females), and age group (all age, <65, or 65+). Incidence rates included all race/ethnicity groups and cancer stages. Descriptive statistics of cancer incidence rates are presented in Table 2. As stated on the State Cancer Profiles database, county cancer data was suppressed if there were less than 16 cases of a given cancer. For the purposes of the current analysis, suppressed county cancer data was treated as missing data.

Table 2. Descriptive statistics for county-level cancer incidence rates per 100,000 population by state

	Iowa			Illinois		
Cancer Site	Sample Size	Mean (SD)	Range	Sample Size	Mean (SD)	Range
Colon and Rectum						
-Overall	97	45.90 (8.71)	29.60-75.60	98	47.10 (7.81)	34.30-67.80
-Male	83	53.00 (12.30)	29.50-99.90	88	53.60 (9.53)	35.10-73.10
-Female	79	41.50 (9.24)	22.90-58.90	88	41.30 (8.35)	27.20-63.80
-<65	53	22.20 (6.76)	12.00-49.50	80	22.30 (4.82)	14.30-43.80
-<65, Male	27	24.40 (6.73)	12.20-42.50	58	26.10 (6.84)	13.20-45.10
-<65, Female	16	18.90 (3.50)	13.30-26.50	46	19.10 (5.45)	10.80-42.20
-65+	90	223.00 (44.60)	132.00-321.00	92	223.00 (48.50)	122.00-392.00
-65+, Male	49	262.00 (60.20)	162.00-425.00	73	255.00 (50.50)	174.00-424.00
-65+, Female	60	207.00 (51.00)	137.00-312.00	70	197.00 (47.70)	115.00-363.00
Lung and Bronchus						
-Overall	99	62.70 (11.70)	29.90-91.80	102	76.30 (15.50)	50.20-134.00
-Male	95	75.90 (16.90)	36.20-128.00	99	91.50 (21.20)	52.70-147.00
-Female	86	54.80 (12.00)	26.20-84.80	96	65.40 (13.30)	45.00-125.00
-<65	63	22.50 (5.48)	11.30-40.90	90	27.60 (8.88)	12.10-67.20
-<65, Male	27	24.00 (6.03)	12.70-40.00	74	29.90 (10.50)	12.40-64.60
-<65, Female	25	23.00 (7.43)	12.90-40.40	60	26.00 (8.32)	11.90-46.20
-65+	97	354.00 (66.80)	159.00-536.00	101	417.00 (78.30)	277.00-643.00
-65+, Male	85	456.00 (107.00)	233.00-777.00	94	518.00 (117.00)	307.00-936.00
-65+, Female	72	305.00 (68.50)	196.00-493.00	90	342.00 (70.90)	204.00-584.00
Breast (Females Only)						
-Overall	98	125.00 (23.60)	69.60-220.00	100	128.00 (18.30)	80.30-173.00
-<65	84	84.30 (17.10)	39.80-155.00	90	85.40 (13.40)	51.70-124.00
-65+	89	443.00 (80.60)	241.00-674.00	92	442.00 (75.10)	278.00-681.00
Prostate (Males Only)						
-Overall	99	109.00 (26.00)	56.70-177.00	100	105.00 (17.60)	66.00-159.00
-<65	74	47.70 (13.40)	27.60-85.40	80	42.70 (8.72)	17.60-67.60
-65+	90	581.00 (124.00)	347.00-875.00	95	544.00 (99.90)	331.00-806.00

2.1.3 Urbanicity Identification

Identification of a county as metropolitan, micropolitan, or rural was based on designation by the US Health Resources and Service Administration's Federal Office of Rural Health Policy (FORHP). The available FORHP dataset contains only those counties in the United States that constitute rural areas as per FORHP guidelines, including metros that are located in a rural census tract and both micropolitan and rural areas. Six counties in Illinois were not linked by the FORHP as they were identified as being part of the Chicago metropolitan area. Instead, they were manually added to the dataset as urban counties based on Core Based Statistical Areas (CBSA) data available on the Rural Health Information Hub website. A summary of CBSA category distributions by state is presented in Table 3.

Table 3. Descriptive statistics for CBSA county types by state

	Iowa		Illinois	
CBSA County Classification	Count	Percent of Counties (%)	Count	Percent of Counties (%)
Metropolitan	21	21.2	40	39.2
Micropolitan	19	19.2	24	23.5
Rural	59	59.6	38	37.3

2.2 Statistical Analysis

All data management and analyses were performed using R. Specifically for the purpose of model building, the "mgcv" package was used to fit generalized additive models while the "spdep" package was used to fit spatial econometrics models.

2.2.1 Gaussian Generalized Additive Model

Exploratory analyses were conducted to examine associations between each contaminant type and the overall incidence rate for each of the four cancers; the incidence rates as stratified by sex or age were excluded from the main analysis with the possibility of further analysis at a later time. Both linear and non-linear associations between incidence rates and contaminants were considered using penalized splines. An F-test was used to compare whether the models with smooth terms were

significantly better than their counterparts with only linear terms at α=0.05. All models had the following form:

$$y_i = \beta_0 + f(CONC_i) + \beta_1 ST_i + \beta_2 CBSA_i + \epsilon_i$$

where for county i, y_i was the overall county-specific incidence rate of a single cancer type (colon and rectum, lung and bronchus, breast, or prostate), $CONC_i$ referred to a given contaminant's county-specific mean concentration (as a mean of all available years or as a mean of a specific decade), and $f(\,)$ denoted either a linear or smooth function. The models were also adjusted for two geographic variables: ST_i (identification of a county as being in Iowa or Illinois) and $CBSA_i$ (classification of a county as being metropolitan, micropolitan, or rural). Model parameter estimates and standard errors for $CONC_i$ were then used to calculate the estimated risk and 95% confidence intervals, per standard deviation for each $CONC_i$ exposure, making the final risk values more comparable between exposures and exposure window years.

2.2.2 Simultaneous Autoregressive Model

The initial generalized additive models did not account for residual spatial dependence between county-level cancer incidence rates. In reality, counties are not separate entities; their populations may frequently move around across several neighboring counties in order to access key resources. Therefore, it is likely that the location of an individual's exposure to a certain drinking water contaminant is not restricted to just that individual's county of residence. In addition, neighboring counties are likely to share similar spatially-varying risk factors. A spatial analysis would thus control for spatial dependence not explained by the model.

We considered a spatial simultaneous autoregressive (SAR) model, which allows us to account for the effects on the outcome variable from a given county as attributable to immediately-

neighboring counties or the neighbors of those immediate neighbors. In this manner, the data was examined using the following simultaneous autoregressive model:

$$Y_i = \beta_0 + \beta_1 CONC_i + \beta_2 ST_i + \beta_3 CBSA_i + \rho \boldsymbol{W_i}(\boldsymbol{Y} - \boldsymbol{\beta X}) + \epsilon_i$$

where Y_i was the overall county-specific incidence rate of a single cancer type (colon and rectum, lung and bronchus, breast, or prostate). However, this model now accounted for $\boldsymbol{W_i}$, or the row-standardized matrix of spatial adjacency effects as first-order (due to immediate neighbors) or second-order (due to neighbors of immediate neighbors). Whether the first-order or second-order effects yielded different results were assessed by comparing the AIC values of these models and to that of the original generalized additive models. From this, the model with the lowest significant AIC value was selected and used to derive the parameter estimates and standard errors for β_1.

3. Results

3.1 Spatial Visualizations

For cancer and 2010-2017 exposure, spatial patterns in contaminant concentration levels and cancer incidence were visualized with choropleth maps.

3.1.1 Distributions of Mean Contaminant Concentrations From 2010-2017

Focusing specifically on the 2010-2017 exposure window, the mean concentration for each contaminant was plotted for each county in Iowa and Illinois, resulting in five separate choropleth maps as indicated in Figure 1.

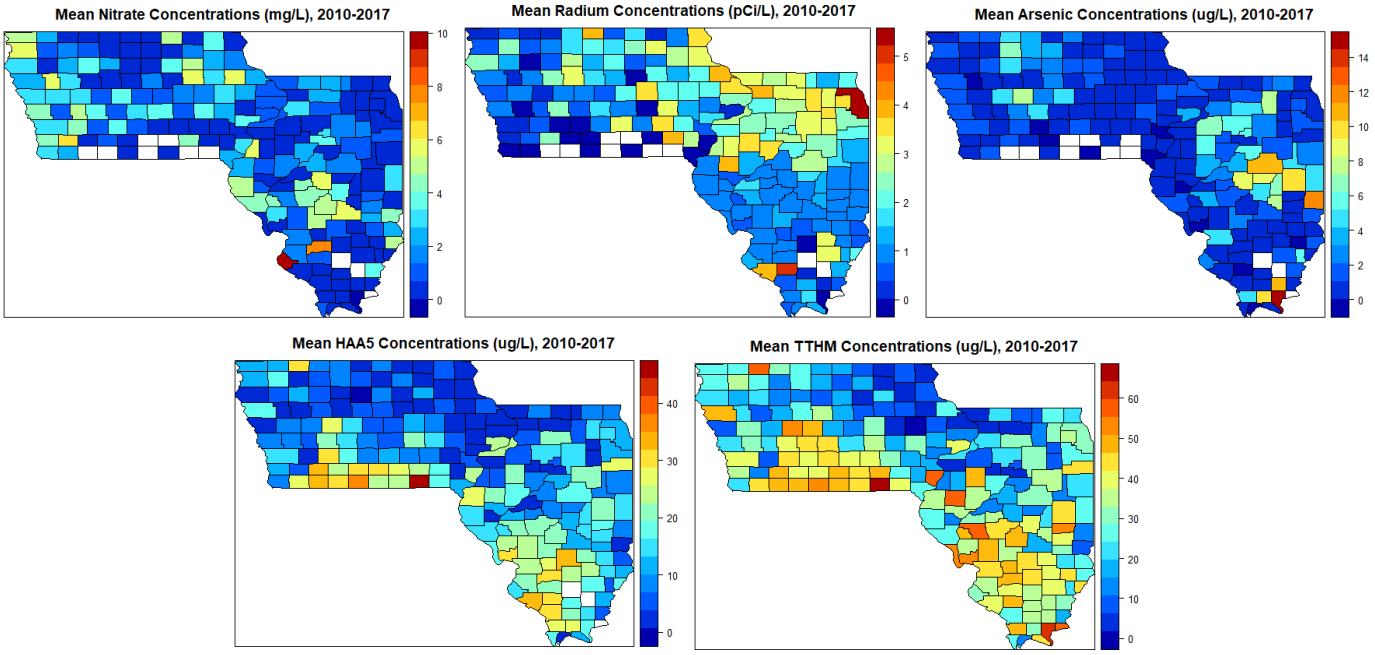

Figure 1. From left-to-right and top-to-bottom, choropleth plots for the mean contaminant concentration during the 2010-2017 exposure window for (a) nitrate, (b) radium, (c) arsenic, (d) haloacetic acids, and (e) total trihalomethanes

Based on the patterns observed in Figure 1, we note that counties in Iowa reported wider ranges of concentrations for radium, haloacetic acid, and total trihalomethanes, but for nitrate and arsenic, most counties with available data reported means that fell in the lower half of the spectra for those contaminants. Counties in Illinois reported wider ranges of concentrations for all contaminants except nitrate and arsenic, where isolated pockets of moderate values appeared in central Illinois. For radium, both states reported wide concentration ranges; values in the upper half of the concentration spectrum (>2.5 pCi/L) occurred especially in far-northeastern Iowa and stretching across northern Illinois. Higher values for haloacetic acids (>20 µg/L) and total trihalomethanes (>35 µg/L) occurred in southern Iowa and southern Illinois, especially towards these states' borders with Missouri.

3.1.2 Distributions of Five-Year Cancer Incidence Rates From 2013-2017

The five-year incidence rates for each cancer site were plotted for each county in Iowa and Illinois, resulting in four separate choropleth maps as indicated in Figure 2.

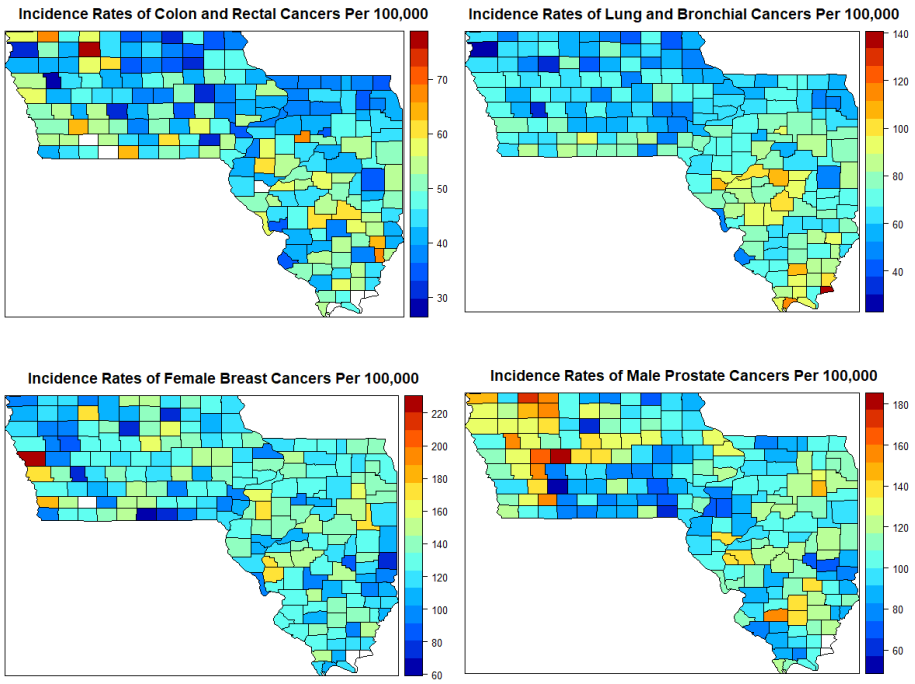

Figure 2. From left-to-right and top-to-bottom, choropleth plots for the five-year incidence rates, 2013-2017, for (a) colon and rectal cancers, (b) lung and bronchial cancers, (c) female breast cancers, and (d) male prostate cancers

Based on the patterns observed in Figure 2, it was noted that for all four cancer sites, there was a wide range of cancer incidence rate values in both Iowa and Illinois. For colon and rectal cancer, female breast cancer, and male prostate cancer, the lowest and highest values for incidence rates tended to occur in Iowa. As a whole, lower rates of colon and rectal cancers were especially present in northeastern Iowa and northern Illinois (30-55 cases per 100,000), but west-central Illinois showed a cluster of higher incidence for lung and bronchial cancer (90-110 cases per 100,000). Apart from the maximum value in western Iowa, the spread of female breast cancer incidence rates was roughly heterogeneous across both states. Western Iowa also tended to have a more distinct clustering of counties with high incidence rates for prostate cancer (>130 cases per 100,000) while similar clusters were smaller and more spread out in northern and southern Illinois.

3.2 Exploratory Analyses

The purpose of performing initial analyses with a GAM was to examine relationships between contaminant exposure and cancer incidence rates while controlling for state (reference of Iowa) and urbanicity classification (reference of metropolitan). As seen in Table 4, the resulting linear effects of this model yielded the following values of change in incidence rates.

Table 4. Estimated changes in mean cancer incidence per 100,000 per standard deviation (SD) increase in contaminant and 95% confidence intervals (CI). Statistically significant associations ($p<0.05$) are bolded and indicated by *.

Contaminant	Exposure Window	SD	Colon Cancer	Lung Cancer	Breast Cancer	Prostate Cancer
Nitrate	1990-2017	1.89	-0.80 (-2.06, 0.45)	0.10 (-1.86, 2.05)	0.66 (-2.51, 3.83)	2.99 (-0.20, 6.18)
	1990-1999	2.01	0.62 (-0.64, 1.87)	0.72 (-1.24, 2.68)	1.04 (-2.13, 4.20)	1.96 (-1.86, 5.77)
	2000-2009	1.70	-0.07 (-1.26, 1.12)	1.01 (-0.84, 2.86)	1.32 (-1.62, 4.26)	-0.07 (-1.26, 1.12)
	2010-2017	1.83	-0.14 (-1.33, 1.06)	-0.01 (-1.87, 1.85)	0.86 (-2.11, 3.82)	2.42 (-0.78, 5.61)
Radium	1990-2017	1.88	-0.44 (-2.07, 1.19)	-2.20 (-4.72, 0.32)	**4.79 (0.77, 8.80) ***	-0.24 (-4.66, 4.18)
	1990-1999	2.68	-0.03 (-2.28, 2.23)	1.86 (-1.18, 4.89)	1.10 (-4.71, 6.90)	1.07 (-5.44, 7.58)
	2000-2009	1.99	**-1.66 (-3.18, -0.14) ***	**-3.13 (-5.48, -0.78) ***	1.75 (-2.10, 5.59)	-0.68 (-4.86, 3.51)
	2010-2017	1.55	**-1.79 (-3.35, -0.23) ***	**-4.64 (-7.01, -2.28) ***	2.50 (-1.41, 6.42)	-4.04 (-8.24, 0.17)
Arsenic	1990-2017	1.83	-0.02 (-0.73, 0.70)	0.22 (-0.86, 1.31)	-1.13 (-2.91, 0.66)	0.31 (-1.59, 2.20)
	1990-1999	1.23	0.07 (-0.56, 0.70)	0.57 (-0.40, 1.54)	-1.07 (-2.64, 0.50)	0.54 (-1.24, 2.32)
	2000-2009	2.21	0.08 (-0.52, 0.69)	0.19 (-0.76, 1.13)	-0.63 (-2.14, 0.88)	0.94 (-0.67, 2.56)
	2010-2017	1.92	0.47 (-0.50, 1.44)	-0.02 (-1.44, 1.40)	-0.47 (-2.90, 1.95)	1.42 (-1.01, 3.86)
HAA5	1990-2017	13.70	0.02 (-1.47, 1.51)	0.80 (-1.49, 3.09)	-0.40 (-4.09, 3.29)	1.96 (-2.04, 5.95)
	1990-1999	19.40	0.73 (-1.22, 2.67)	-0.28 (-3.64, 3.09)	1.28 (-2.65, 5.20)	-2.50 (-6.50, 1.50)
	2000-2009	14.30	0.98 (-0.34, 2.30)	**2.83 (1.01, 4.66) ***	0.02 (-3.00, 3.04)	-1.78 (-5.03, 1.47)
	2010-2017	12.80	1.04 (-0.58, 2.67)	2.27 (-0.18, 4.73)	0.10 (-3.93, 4.13)	-2.34 (-6.65, 1.97)
TTHM	1990-2017	30.30	0.75 (-0.93, 2.43)	**2.51 (0.92, 4.10) ***	-0.82 (-3.34, 1.70)	2.22 (-0.50, 4.94)
	1990-1999	24.50	**1.59 (0.03, 3.14) ***	**2.56 (0.13, 4.99) ***	0.78 (-2.91, 4.46)	1.36 (-2.71, 5.44)
	2000-2009	35.50	**1.61 (0.20, 3.03) ***	**1.47 (0.62, 2.31) ***	-0.58 (-1.92, 0.76)	0.42 (-1.03, 1.87)
	2010-2017	27.70	**2.42 (0.27, 4.58) ***	**4.32 (0.96, 7.68) ***	-1.21 (-6.65, 4.23)	1.83 (-3.97, 7.62)

Radium was found to be significantly associated with cancer incidence counts (illustrated by the estimated changes in Table 4). Focusing just on colon and rectal cancers, significant negative associations were found with radium concentrations during two of the three decade-specific exposure windows (Table 4), where there was a decrease in colon and rectal cancers of 1.66 cases per 100,000 (95% CI: -3.18, -0.14) for exposures during 2000-2009 and of 1.79 cases per 100,000 (95% CI: -3.35, -0.23) for exposures during 2010-2017. These two same exposure windows also yielded significant negative associations with lung and bronchial cancers (Table 4), where there was a decrease of 3.13 cases per 100,000 (95% CI: -5.48, -0.78) and 4.64 cases per 100,000 (95% CI: -7.01, -2.28) for exposures during 2000-2009 and during 2010-2017, respectively. On the contrary, radium was positively associated with the incidence of female breast cancer, where there was an increase of 4.79 cases per 100,000 (95% CI: 0.77, 8.80) for the full 1990-2017 exposure window.

The by-products of water chlorination were also found to be significantly associated with cancer incidence counts. Haloacetic acids yielded a significant positive association with lung and bronchial cancers during only one of the three decade-specific exposure windows (Table 4), where there was an increase in lung and bronchial cancers of 2.83 cases per 100,000 (95% CI: 1.01, 4.66) for exposures during 2000-2009. Total trihalomethanes yielded positive significant associations with colon and rectal cancers during all three decade-specific exposure windows (Table 4), where there was an increase of 1.59 cases per 100,000 (95% CI: 0.03, 3.14) for exposures during 1990-1999, of 1.61 cases per 100,000 (95% CI: 0.20, 3.03) for exposures during 2000-2009, and of 2.42 cases per 100,000 (95% CI: 0.27, 4.58) for exposures during 2010-2017. Total trihalomethanes also yielded positive significant associations for lung and bronchial cancers during the full timeframe and during all three decade-specific exposure windows (Table 4), where there was an increase of 2.56 cases per 100,000 (95% CI: 0.13, 4.99) for exposures during 1990-1999, of 1.47 cases per 100,000 (95% CI:

0.62, 2.31) for exposures during 2000-2009, of 4.32 cases per 100,000 (95% CI: 0.96, 7.68) for exposures during 2010-2017, and of 2.51 cases per 100,000 (95% CI: 0.92, 4.10) for the full 1990-2017 exposure window.

None of the contaminant exposure windows were significantly associated with prostate cancer incidence (Table 4). Likewise, none of the other contaminants (nitrate, arsenic) yielded significant associations with colon and rectal, lung and bronchial, or female breast cancers.

Non-linear effects were also considered from this model. While several contaminant exposure windows had non-linear associations with cancer incidence, only five non-linear associations were statistically significant at $\alpha=0.05$. Radium concentrations from the 2000-2009 exposure window yielded a significant non-linear association with lung and bronchial cancer incidence ($F=3.499$, $p=0.008$) (Figure 3).

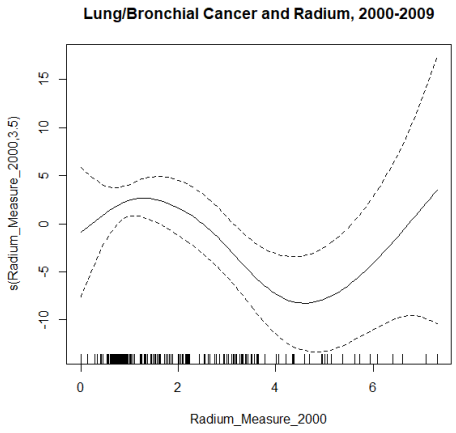

Figure 3. Plot of non-linear effects of 2000-2009 mean radium concentrations on lung and bronchial cancer incidence rates

At α=0.05. haloacetic acid concentrations from the 2000-2009 exposure window also yielded a significant non-linear association with lung and bronchial cancer incidence (F=4.351, p=0.026) (Figure 4).

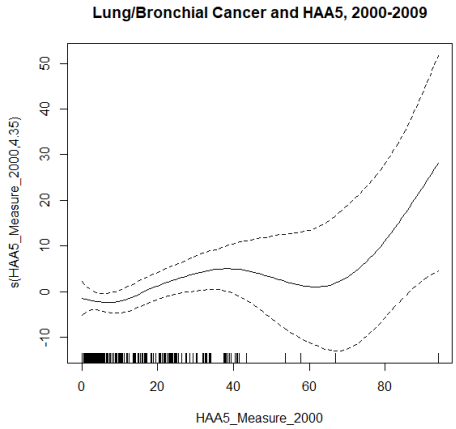

Figure 4. Plot of non-linear effects of 2000-2009 mean haloacetic acid concentrations on lung and bronchial cancer incidence rates

Total trihalomethanes yielded three significant non-linear associations. The 1990-1999 exposure window was associated with lung and bronchial cancer incidence (F=2.327, p=0.019) and with prostate cancer incidence (F=3.419, p=0.0141) (Figure 5), and the 2000-2009 exposure window was associated with only lung and bronchial cancer incidence (F=1.684, p=0.001) (Figure 5). Beyond these findings, non-linear associations were not examined further in the current analysis.

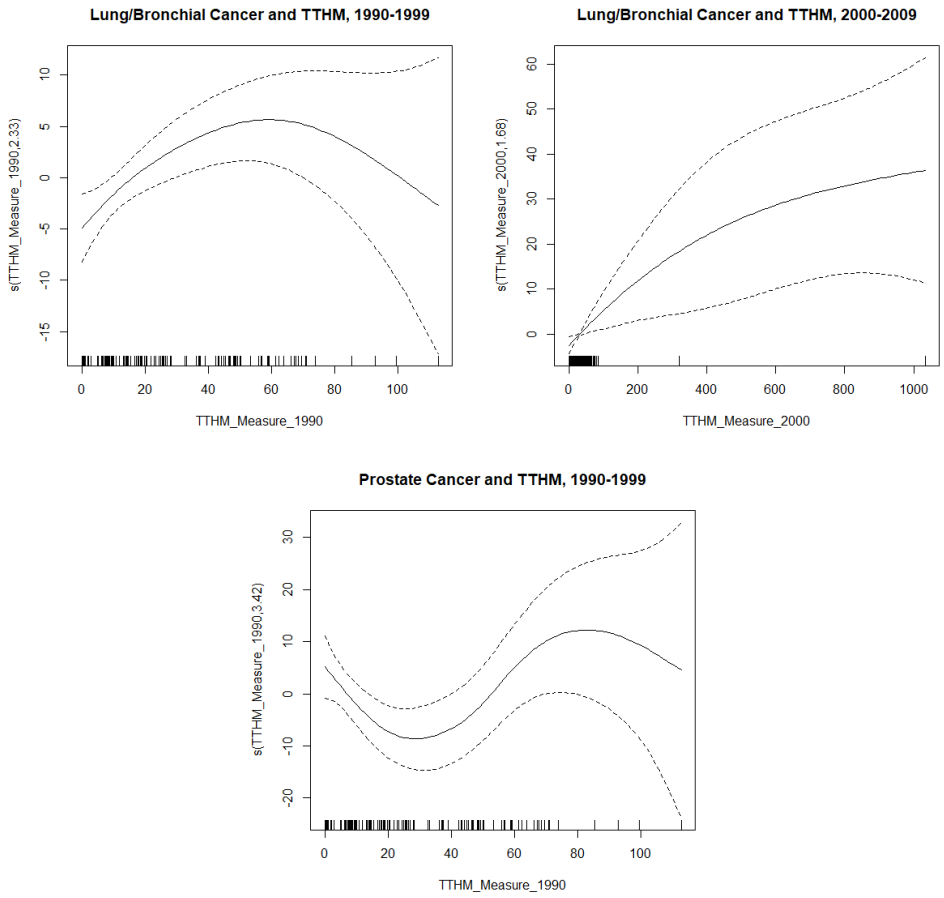

Figure 5. From left-to-right and top-to-bottom, plot of non-linear effects of (a) 1990-1999 mean total trihalomethane concentrations on lung and bronchial cancer incidence rates, (b) 2000-2009 mean total trihalomethane concentrations on lung and bronchial cancer incidence rates, and (c) 1990-1999 mean total trihalomethane concentrations on prostate cancer incidence rates

3.3 SAR Modeling

The SAR model identified relationships between contaminant exposure and cancer incidence rates while controlling for state and urbanicity classification and considering spatial dependence. As seen in Tables 5 and 6, the results of this model allowed for selection between first-order and second-order spatial dependence and their resulting values of change in cancer incidence rates. For simplicity, the exposure windows were limited to those from the 2010-2017 timeframe.

Table 5. Spatial dependence estimates and parameters for model comparisons. Statistically significant associations ($p<0.05$) are bolded and indicated by *.

		Colon			Lung			Breast			Prostate		
Contaminant	**Weight**	**λ**	**σ^2**	**AIC**	**λ**	**σ^2**	**AIC**	**λ**	**σ^2**	**AIC**	**λ**	**σ^2**	**AIC**
Nitrate	1st-order	0.07	65	1325	-0.03	159	1525	0.19	393	1681	-0.20	458	1719
	2nd-order	0.17	64	1324	-0.24	157	1523	0.02	401	1683	-0.00	468	1721
Radium	1st-order	0.06	63	1320	0.02	148	1510	0.19	390	1679	-0.14	460	1719
	2nd-order	0.16	63	1319	-0.33	143	1507	-0.00	398	1682	-0.00	465	1720
Arsenic	1st-order	0.08	64	1315	0.01	159	1516	0.19	395	1673	-0.14	459	1710
	2nd-order	0.12	64	1315	-0.26	156	1514	0.02	403	1675	0.02	464	1711
HAA5	1st-order	0.08	64	1364	-0.05	157	1577	0.17	407	1749	-0.14	478	1789
	2nd-order	0.17	64	1364	-0.25	155	1575	-0.04	413	1751	0.02	483	1790
TTHM	1st-order	0.10	62	1374	0.04	167	1613	0.16	406	1766	-0.11	482	1809
	2nd-order	0.15	62	1374	**-0.33**	**162**	**1609***	-0.05	411	1768	-0.01	485	1810

Table 6. Estimated changes in mean cancer incidence per 100,000 per standard deviation (SD) increase in contaminant and 95% confidence intervals (CI), accounting for spatial dependence. Statistically significant associations ($p<0.05$) are bolded and indicated by *.

Contaminant	**SD**	**Colon Cancer**	**Lung Cancer**	**Breast Cancer**	**Prostate Cancer**
Nitrate	1.83	-0.14 (-1.32, 1.03)	0.17 (-1.66, 2.00)	1.22 (-1.66, 4.11)	2.78 (-0.33, 5.89)
Radium	1.55	**-1.75 (-3.28, -0.23) ***	**-4.83 (-7.10, -2.56) ***	2.74 (-0.98, 6.46)	-3.79 (-7.98, 0.40)
Arsenic	1.92	0.51 (-0.44, 1.47)	-0.05 (-1.46, 1.37)	-0.62 (-2.98, 1.73)	1.20 (-1.20, 3.59)
HAA5	12.80	1.02 (-0.58, 2.62)	2.30 (-0.11, 4.70)	0.08 (-3.87, 4.02)	-2.84 (-7.05, 1.38)
TTHM	27.70	**2.43 (0.31, 4.54) ***	**4.14 (0.90, 7.38) ***	-0.87 (-6.18, 4.43)	1.23 (-4.48, 6.93)

We selected the type of spatial dependency by comparing the AIC values between the first and second-order models; a lower AIC value indicated better model fit. First-order spatial dependency was selected unless the AIC for the second-order model was at least two integer values smaller than the AIC of the first-order model. As shown in Table 5, for colon and rectal cancer, first-order spatial dependency provided better fit for each of the five possible contaminant types based on the AIC

values. For lung and bronchial cancer, second-order spatial dependency was preferred for each of the five contaminant types. All five contaminant types also demonstrated first-order spatial dependency when modeled with female breast and prostate cancers. At $\alpha=0.05$, only the second-order value for lambda (λ) in the model of total trihalomethanes and lung and bronchial cancers was statistically significant ($p=0.049$), thus indicating that this association has spatial dependence based on the effects of the neighbors of neighboring counties. No other values for lambda were statistically significant, thus indicating that the other associations between contaminants and cancer incidence are more dependent on immediate county effects rather than on the effects of neighboring counties (Table 5).

Radium concentrations yielded a significant negative association with both colon and rectal cancer incidence and with lung and bronchial cancer incidence (Table 6). Each 1.55 pCi/L increase in a county's mean drinking water radium concentration from 2010-2017 decreased that county's incidence rate of colon and rectal cancers by 1.75 cases per 100,000 (95% CI: -3.28, -0.23) and the incidence rate of lung and bronchial cancers by 4.83 cases per 100,000 (95% CI: -7.10, -2.56). These significant associations were similar to the negative associations found with radium as observed from GAM, but the magnitude of these estimates were not significantly different between GAM and SAR. Total trihalomethanes also yielded significant positive associations with both colon and rectal cancer incidence and with lung and bronchial cancer incidence (Table 6), where there was an increase of 2.43 cases per 100,000 (95% CI: 0.31, 4.54) for colon and rectal cancers and of 4.14 cases per 100,000 (95% CI: 0.90, 7.38) for lung and bronchial cancers. These significant associations were also similar to the positive associations found with total trihalomethanes as observed from GAM, but the magnitude of these estimates were not significantly different between GAM and SAR. No other contaminants were significantly associated with colon and rectal cancer or with lung and bronchial cancer. In addition, none of the five contaminants at the 2010-2017 exposure window were

significantly associated with either female breast cancer or prostate cancer, reflecting a similar finding as in the GAM model.

4. Discussion and Conclusions

From the spatial regression analysis, it was found that mean radium concentrations from 1990-2017 were positively associated with the incidence of female breast cancer. This association supported the past findings of Bean et al. (1982), where radium-226 concentrations above the EPA legal limit of 5 pCi/L were positively associated with female breast cancer risk. This is an interesting finding in that despite the gap in time between the current analysis and Bean et al. (1982), the positive association between radium and female breast cancer was observed. Bean et al. (1982) has indicated that water softeners greatly reduce radium content in drinking water. Therefore, it may be beneficial to further study this positive relationship with respect to whether the use of water softeners reduces contaminant content in drinking water.

Radium concentrations were negatively associated with the incidence of lung and bronchial cancer. This differed from Bean et al. (1982), who instead found a positive association among males. The difference between these findings raises the question as to whether stratifying cancer incidence by sex would change the direction of association that we found. By not examining the role of sex in cancer incidence, there is a risk that the overall lung and bronchial cancer incidence values mask sex-stratified trends.

We also found that radium concentrations were negatively associated with the incidence of colon and rectal cancer, but no past studies could be found that examined these associations. Past literature instead focused on the role of nitrate concentrations (not radium). We did not find any significant associations between nitrate and colon and rectal cancer. However, the direction of this association has varied between studies; Weyer et al. (2001) found nitrate concentrations to be

negatively associated with rectal cancers while Schullehner et al. (2018) found the association with colorectal cancers to be positive. The differences in findings could be due to both studies using different methods to define the cancer outcomes. Weyer et al. (2001) used an approach that was more similar to the current analysis: cancer data limited to the years 1986-1998 were taken from the Iowa Cancer Registry. Schullehner et al. (2018) instead identified cancer cases through participant follow-up.

Like with radium, we also found that the two groups of chlorination byproducts (haloacetic acids and total trihalomethanes) were significantly associated with cancer incidence. In the case of colon and rectal cancer, total trihalomethanes were positively associated with cancer incidence, supporting the past findings of Dobaradaran et al. (2019), who found that bromoform increased the incremental lifetime risk for colon and rectal cancer. However, as the current analysis only considered total trihalomethane concentrations, we cannot implicate bromoform for this effect unless we develop models that examine each trihalomethane individually.

In the case of lung and bronchial cancers, both haloacetic acids and total trihalomethanes were positively associated with cancer incidence. This does not support the findings of Morris (1995), who found that chlorination byproducts posed no significant risk for colon or lung cancers but instead increased the risk of rectal cancer (which is instead partly supported by the previously-described findings for colon and rectal cancer). However, as Morris (1995) conducted a meta-analysis, differences in findings could be due to (1) which haloacetic acids and trihalomethanes were included in the meta-analysis and (2) that the current analysis only included both total-group variables, potentially masking the effects of individual haloacetic acid or trihalomethane types.

We found no contaminants to be significantly associated with the incidence of prostate cancer, in contradiction to Roh et al. (2017), who found a positive association between arsenic and prostate

cancer risk. This difference in findings could possibly be explained by how arsenic concentration was defined in each study design. We treated arsenic concentrations as continuous values, but Roh et al. (2017) defined arsenic concentrations as tertiles of concentration level. The range of mean values used in the current analysis overlapped the values defining the tertile ranges, but these tertiles also did not include any values of zero concentration. Therefore, it may be valuable to further consider categorizing mean concentration values or limiting mean concentrations to nonzero values.

For contaminants during the 2010-2017 exposure window, several patterns in contaminant concentrations and cancer incidence were found. For example, a commonality between both states was the variation in radium, which tended to be more concentrated along the Iowa-Illinois border and into northern Illinois. This finding regarding radium is supported by Szabo et al. (2012); the Mid-Continent and Ozark Plateau Cambro-Ordovician aquifer system, which has been identified as having some of the highest radium readings for drinking water in the United States, stretches across most of Iowa and into northern and central Illinois. This aquifer system may play a role regarding the clustering of higher radium concentrations in the current analysis.

The current analysis indicated lower arsenic concentrations for counties in northeast Iowa and northwest Illinois, which does not agree with the findings of Szabo et al. (2012). The Late Wisconsinan glacial drift covered an area that presently includes far-northeastern Iowa and far-northwestern Illinois, and these same areas have previously recorded higher groundwater arsenic levels (Erickson & Barnes, 2005; Szabo et al., 2012).

Similarly, while the current analysis found higher concentration values for haloacetic acids and total trihalomethanes in southern Iowa and southern Illinois, no past studies were identified as having directly examined this spatial pattern. However, as discussed by Allaire et al. (2018), rural low-income counties tend to report more violations of chlorination by-product limits compared to

urban or higher-income counties. It may be useful to further study where rural counties are located in the two states and whether these counties have significantly higher mean concentrations of haloacetic acids or total trihalomethanes.

Based on the mapping of cancer incidence rates, it was found that for all four cancer sites, both Iowa and Illinois reported a wide range of incidence rates. Of the clusters present in the four maps, the larger cluster of higher prostate cancer incidence in western Iowa was of particular interest. This cluster pattern was also observed by Mandal et al. (2009), where northwestern Iowa in particular was part of a larger cluster of significantly-high prostate cancer incidence rates. These authors proposed that higher prostate cancer rates could be related to Vitamin D deficiencies, especially as this region sees lower ultraviolet exposure during the winter.

The current analysis considered spatial dependence in the relationships between contaminant exposure and cancer incidence. Spatial dependence may be present in county-level incidence rates due to other similar cancer risk factors including radon exposure (via inhalation), air quality, or smoking. On the other hand, there were a few shortcomings in the current analysis methods that could be improved upon in future replications or extensions of these methods. The first improvement would be to include more socio-demographic data. For example, data from the Behavioral Risk Factor Surveillance System (BRFSS) from the Centers for Disease Control and Prevention may be useful for this purpose. Raw BRFSS data could be obtained directly from the CDC website and matched to the item names as provided in the measure guides, but a more straightforward approach would be to download summarized state BRFSS data from state cancer or health agency websites. Likewise, county-level radon values from Iowa and Illinois public-use datasets could be included as an additional covariate; Stanley et al. (2019) stated that radon exposure is the leading cause for lung cancer among non-smokers and is the second most-likely cause among smokers. Including radon

would enable us to specifically control for radon levels when studying the associations between contaminant exposure and cancer incidence.

A recommended future direction would be to further extend the current SAR model to include these additional radon, behavioral, and socio-demographic covariates. In addition, the grouping of contaminant data by exposure windows could be modified to preserve data to each year in the total time-frame or to use smaller multi-year windows. Lastly, future analyses could also be expanded to include data from Minnesota and Wisconsin. Given the geographical proximity of these states with Iowa and Illinois, it might be useful to examine the relationships between contaminants and cancer incidences in this region at-large.

From the findings, we can identify several implications and suggestions for further analysis. The current models enable the identification of relationships between contaminant exposure and cancer incidence while controlling for state and urbanicity classifications. However, first or second-order spatial dependence is not significant except in the association between total trihalomethane concentrations and lung and bronchial cancer incidence (second-order). The models can be improved further by adding additional variables pertaining to socio-demographic or risk factor data (e.g., sex, smoking prevalence), and analyses can be made more specific by using cancer incidence data as limited to a given socio-demographic group. Lastly, the model design allows for data from more states to be included; this can enable the current methods to be expanded to the larger Upper Midwest.

References

Allaire, M., Wu, H., & Lall, U. (2018). National trends in drinking water quality violations. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(9), 2078–2083. https://doi.org/10.1073/pnas.1719805115

Bean, J. A., Isacson, P., Hahne, R. M., & Kohler, J. (1982). Drinking water and cancer incidence in Iowa. II. Radioactivity in drinking water. *American Journal Of Epidemiology*, *116*(6), 924–932. https://doi.org/10.1093/oxfordjournals.aje.a113494

Blair, A., & Freeman, L. B. (2009). Epidemiologic studies in agricultural populations: Observations and future directions. *Journal of Agromedicine*, *14*(2), 125–131. https://doi.org/10.1080/10599240902779436

Dobaradaran, S., Shabankareh Fard, E., Tekle-Röttering, A., Keshtkar, M., Karbasdehi, V. N., Abtahi, M., Gholamnia, R., & Saeedi, R. (2020). Age-sex specific and cause-specific health risk and burden of disease induced by exposure to trihalomethanes (THMs) and haloacetic acids (HAAs) from drinking water: An assessment in four urban communities of Bushehr Province, Iran, 2017. *Environmental Research*, *182*. https://doi.org/10.1016/j.envres.2019.109062

Erickson, M.L., & Barnes, R.J. (2005). Glacial sediment causing regional-scale elevated arsenic in drinking water. *Groundwater*, *43*(6), 796-805. https://doi.org/10.1111/j.1745-6584.2005.00053.x

Evans, S., Campbell, C., & Naidenko, O. V. (2020). Analysis of cumulative cancer risk associated with disinfection byproducts in United States drinking water. *International Journal of Environmental Research and Public Health*, *17*(6). https://doi.org/10.3390/ijerph17062149

Font-Ribera, L., Gràcia-Lavedan, E., Aragonés, N., Pérez-Gómez, B., Pollán, M., Amiano, P., Jiménez-Zabala, A., Castaño-Vinyals, G., Roca-Barceló, A., Ardanaz, E., Burgui, R., Molina, A. J., Fernández-Villa, T., Gómez-Acebo, I., Dierssen-Sotos, T., Moreno, V., Fernandez-Tardon, G., Peiró, R., Kogevinas, M., & Villanueva, C. M. (2018). Long-term exposure to trihalomethanes in drinking water and breast cancer in the Spanish multicase-control study on cancer (MCC-SPAIN). *Environment International*, *112*, 227–234. https://doi.org/10.1016/j.envint.2017.12.031

Gopal, K., Tripathy, S. S., Bersillon, J. L., & Dubey, S. P. (2007). Chlorination byproducts, their toxicodynamics and removal from drinking water. *Journal of Hazardous Materials*, *140*(1-2), 1-6. https://doi.org/10.1016/j.jhazmat.2006.10.063

Jones, C. S., Nielsen, J. K., Schilling, K. E., & Weber, L. J. (2018). Iowa stream nitrate and the Gulf of Mexico. *PLOS One*, *13*(4). https://doi.org/10.1371/journal.pone.0195930

Manassaram, D. M., Backer, L. C., & Moll, D. M. (2006). A review of nitrates in drinking water: Maternal exposure and adverse reproductive and developmental outcomes. *Environmental Health Perspectives*, *114*(3), 320-327. https://doi.org/10.1289/ehp.8407

Mandal, R., St-Hilaire, S., Kie, J. G., & Derryberry, D. (2009). Spatial trends of breast and prostate cancers in the United States between 2000 and 2005. *International Journal of Health Geographics*, *8*(53), 1-10. https://doi.org/10.1186/1476-072X-8-53

Morris, R. D. (1995). Drinking water and cancer. *Environmental Health Perspectives*, *103 Suppl 8*(Suppl 8), 225–231. https://doi.org/10.1289/ehp.95103s8225

Roh, T., Lynch, C. F., Weyer, P., Wang, K., Kelly, K. M., & Ludewig, G. (2017). Low-level arsenic exposure from drinking water is associated with prostate cancer in Iowa. *Environmental Research*, *159*, 338–343. https://doi.org/10.1016/j.envres.2017.08.026

Rushton, G., Peleg, I., Banerjee, A., Smith, G., & West, M. (2004). Analyzing geographic patterns of disease incidence: Rates of late-stage colorectal cancer in Iowa. *Journal of Medical Systems, 28*(3), 223–236. https://doi.org/10.1023/b:joms.0000032841.39701.36

Schullehner, J., Hansen, B., Thygesen, M., Pedersen, C.B. & Sigsgaard, T. (2018), Nitrate in drinking water and colorectal cancer risk: A nationwide population-based cohort study. *International Journal of Cancer, 143*(1), 73-79. https://doi.org/10.1002/ijc.31306

Siegel, R. L., Miller, K. D., Fuchs, H., & Jemal, A. (2021). Cancer statistics, 2021. *CA: A Cancer Journal for Clinicians, 71*(1), 7-33. https://doi.org/10.3322/caac.21654

Stanley, F. K. T., Irvine, J. L., Jacques, W. R., Salgia, S. R., Innes, D. G., Winquist, B. D., Torr, D., Brenner, D. R., & Goodarzi, A. A. (2019). Radon exposure is rising steadily within the modern North American residential environment, and is increasingly uniform across seasons. *Scientific Reports, 9*(18472). https://doi.org/10.1038/s41598-019-54891-8

Szabo, Z., DePaul, V. T., Fischer, J., M., Kraemer, T. F., & Jacobsen, E. (2012). Occurrence and geochemistry of radium in water from principal drinking-water aquifer systems of the United States. *Applied Geochemistry, 27*(3), 729-752. https://doi.org/10.1016/j.apgeochem.2011.11.002

Weyer, P. J., Cerhan, J. R., Kross, B. C., Hallberg, G. R., Kantamneni, J., Breuer, G., Jones, M. P., Zheng, W., & Lynch, C. F. (2001). Municipal drinking water nitrate level and cancer risk in older women: The Iowa Women's Health Study. *Epidemiology, 12*(3), 327–338. https://doi.org/10.1097/00001648-200105000-00013