



COLLEGE OF PUBLIC HEALTH
The University of Georgia
1785

Applications of Geographic Information Systems in Environmental Epidemiology

John E. Vena, Ph.D., Sara E. Wagner, Ph.D., MSPH and Stephen L. Rathbun
jvena@uga.edu swagner@uga.edu rathbun@uga.edu

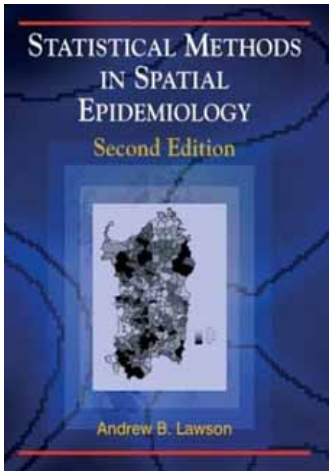
Be Part of the Solution

Outline of the Presentation

- Research Background
- TSP Air pollution and Breast Cancer
- A Study on the Clustering of Residence and Breast Cancer Risk
- DBP in water and bladder and rectal cancer
- Hierarchical Bayesian modeling of the spatio-temporal patterns of lung cancer incidence risk
- Mortality to Incidence Ratios MIR for Cancer
- Statistical Challenges to Linking Spatial Pattern of Cancer to Radiation Exposure
- Discussion

Spatial Epidemiology

- The analysis of spatial/geographical distribution of the incidence of disease
- Spatial hypotheses
  - Disease mapping
  - Ecological analysis
  - Disease clustering
    - General
    - Focused



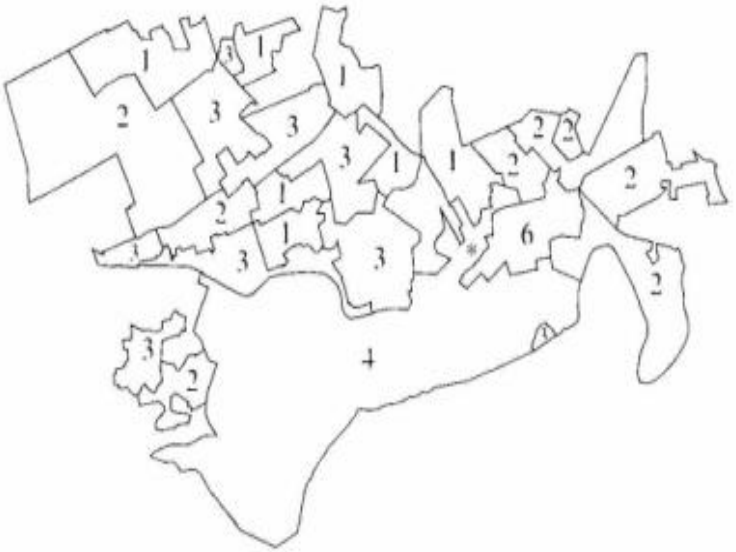
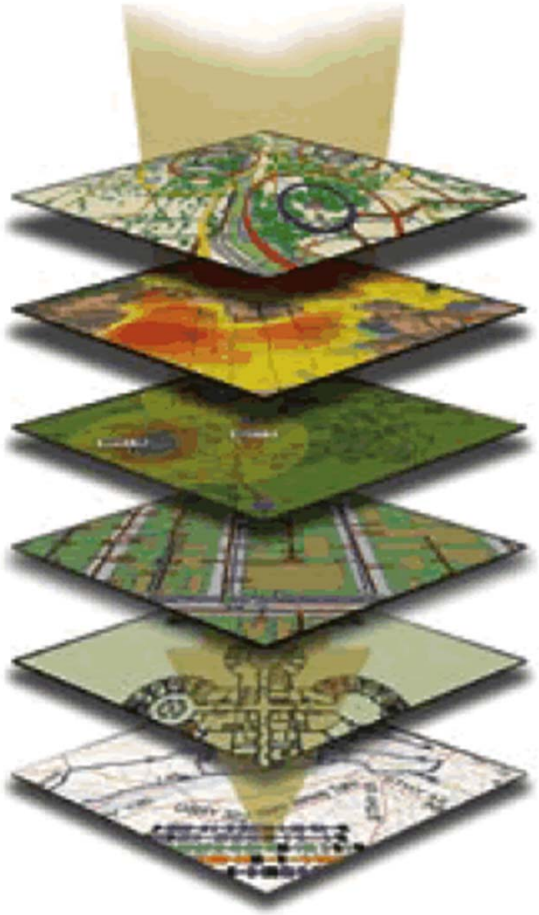


Figure 1.1 Falkirk: central Scotland respiratory cancer counts in 26 census enumeration districts over a fixed time period. * Putative health hazard.

What is a GIS?

- System that captures, stores, analyzes, manages, & presents data linked to location
- Allows: interactive queries, analysis of spatial information, data editing, map creation, result presentation



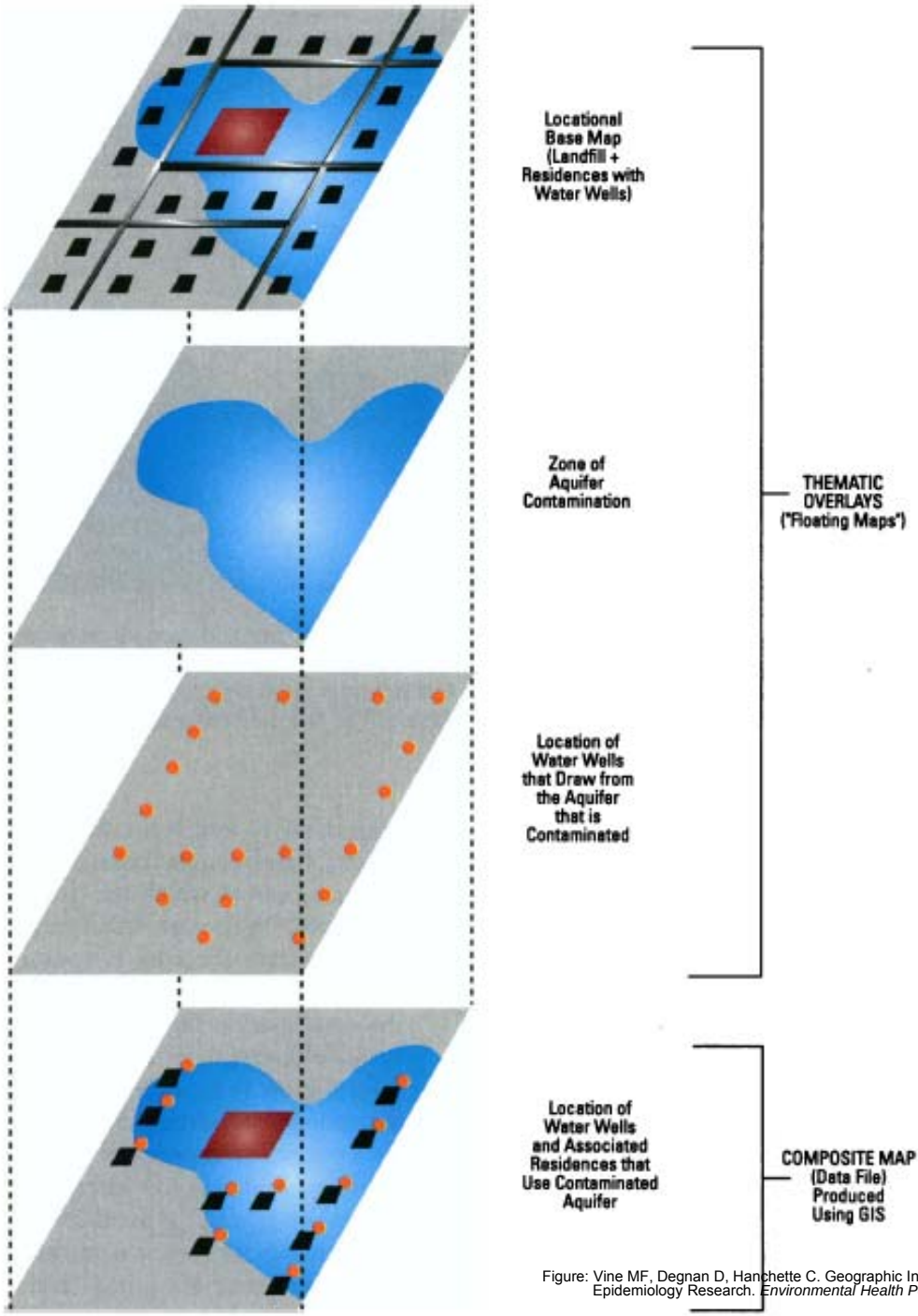
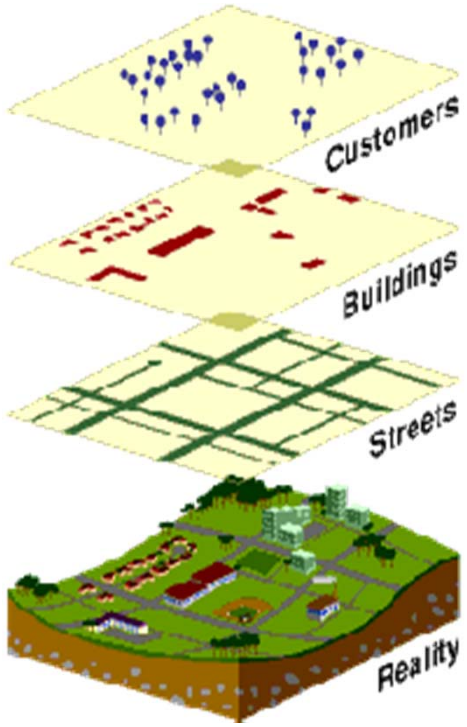


Figure: Vine MF, Degnan D, Hanchette C. Geographic Information Systems: Their Use in Environmental Epidemiology Research. *Environmental Health Perspectives* 1997; 105(6) : 598-605

What is GIS ?

- Combining data from various sources
- Linking multiple databases
- Visualizing data effectively
- Turing data into information: spatial analysis
- “Interactive” maps and databases: Query



Benefits of GIS

- GIS are useful in handling and manipulating large and various sources of datasets.
- GIS are useful for the analysis of different types of data using spatial analysis methods.
- GIS are useful for further analysis with integration of other data or tools.
- GIS are useful for mapping at various geographic scales.

GIS Applications in Public Health

- Disease mapping;

  - Mapping populations at risk

  - Determining spatial patterns of diseases

  - Analyzing spatial and temporal trends

  - Visualizing areas of elevated risk.

- Analytical spatial analyses;

  - Modeling exposures to environmental factors

  - Stratifying risk factors

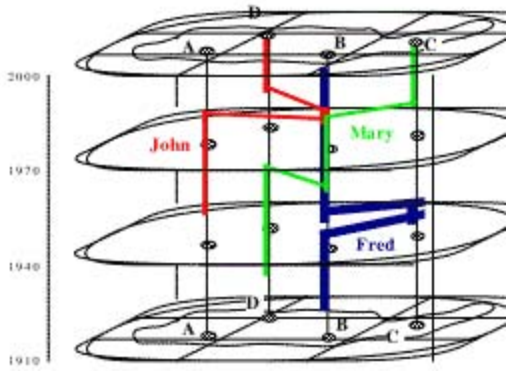
GIS Applications in Public Health

- Detecting disease clustering

- Evaluating health care access and delivery of health services.

- Intervention and prevention;
  - Infectious disease surveillance and control
  - Outbreak investigation and response
  - Assessing resource allocation
  - Planning and targeting interventions
  - Monitoring diseases and interventions over time.

Spatio-Temporal GIS Analysis for Environmental Health



Funding Agency:

National Institute of Environmental Health Sciences

Principal Investigators:

- David Mark
- Max Egenhofer

Collaborators

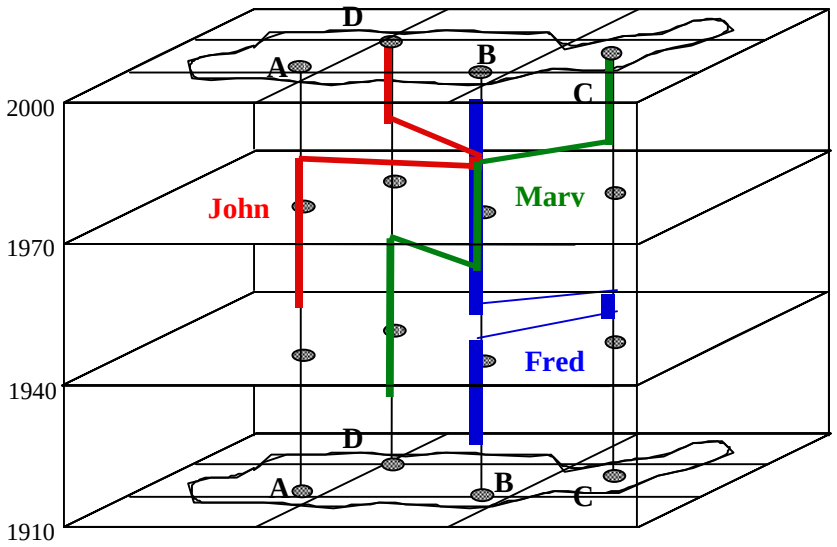
- Ling Bian
- Kathleen Hornsby
- Peter Rogerson
- John Vena

Project Summary:

This research project focuses on the extraction of health-related information from geospatial lifelines, which capture individuals' locations in geographic space a regular or irregular temporal intervals. The objectives of our work are to develop and test the theory of geospatial lifelines in the environmental health sciences

- developing methods to trace locations of individual people (patients, cases, or controls) back through time, to discover spatial clusters in the past or to determine past environmental exposures,
- designing, prototyping, and assess computational models that can deal with large sets of geospatial lifelines and environmental information, and
- examining the ethical and legal implications of recording individuals' geospatial lifelines in databases, and establishing procedures for appropriate restricti

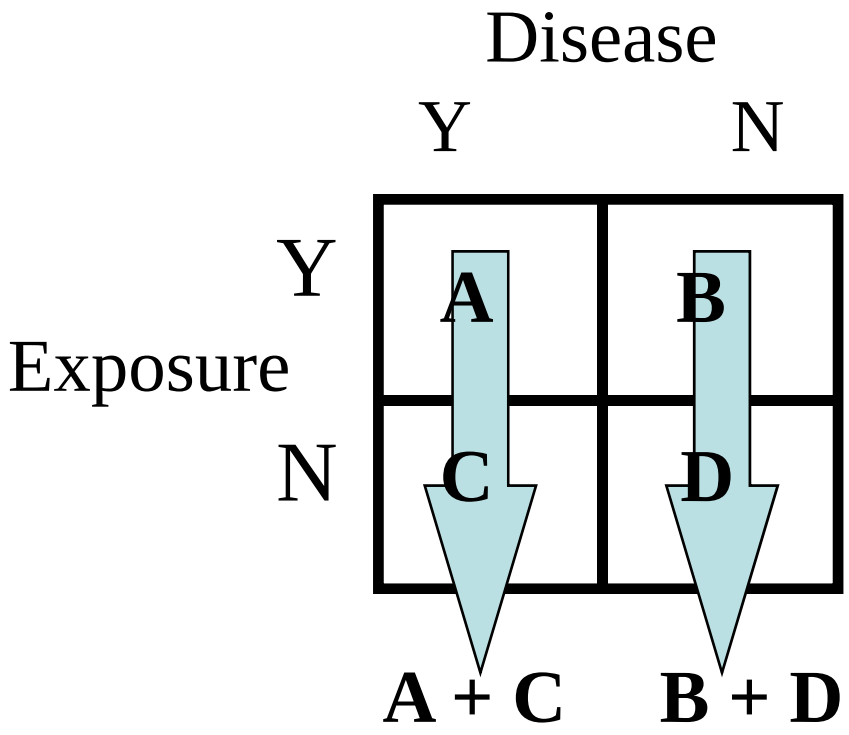
Geospatial Lifelines



Geographical and Environmental Risk Factors for Breast Cancer

- We do not fully understand mechanisms for the known risk factors; eg. Why changes in age at menarche have an impact on breast cancer risk
- There is a substantial geographical variation in breast cancer incidence and mortality in the US (Lacey et al. 2002)
- Environmental risk factors are believed to be involved in breast cancer incidence (Wolff et al. 1996; Laden and Hunter, 1998)

Retrospective (case-control) study



- **Select on disease status**
- **Explore exposure in past**

- **Common exposures**
- **Rare diseases**
- **Moderate numbers**

- **Recall and selection bias**
- **Adequate controls hard to define or obtain**

Western New York Exposures and Breast Cancer (WEB) Study

- Women, age 35-79 with incident, primary, pathologically confirmed breast cancer diagnosed in Erie and Niagara counties during the period 1996-2001.
- Controls were randomly selected and frequency matched to cases on age, race and county of current residence; controls under 65 years of age were selected from a NYSDMV list and those 65 years and over from a HCFA list.

WEB Study

- No previous cancer diagnosis other than non-melanoma skin cancer.
- Extensive in-person interviews and self-administered questionnaires were used to ascertain medical history, diet, lifetime alcohol consumption, residential history, occupational history, and smoking history.

Residence as a Proxy for Exposures

- Based on the life-course approach, residences were used as a proxy for exposures to investigate the relationships between exposures and breast cancer risk.

Lifetime Residential History

- Lifetime residential history of all participants;

  - 20,240 lifetime addresses

  - An average of six addresses for each individual.

- Temporal groups;

  - Residence at birth

  - Residence at menarche, and at women's first birth

  - 20 years & 10 years prior to diagnosis/interview

  - Current residence

Address Matching (Geocoding)

- Process of linking records in two databases
  (eg. place residential location on street map)
- Essential for further spatial analyses
- Steps: preparation, geocoding, and review and evaluation stages.

Address Matching: Steps

- Preparation stage: error checking and standardizing address components of residential history data (event theme).
- Geocoding stage: batch and interactive matching of event theme on reference theme (street map)
- Evaluation stage: review of unmatched addresses and polk searches for incomplete addresses.

Geocoding Results

Geocoding of residential history	Count (%)
Lifetime residential history obtained Erie and Niagara county addresses Addresses for six temporal groups	20,240 15,487 14,493 (100%)
Total matched a) Batch match b) Interactive match	13,405 (92.5%) 12,404 (85.6%) 1,001 (6.9%)
Total unmatched a) Unable to match b) Incomplete addresses	1,088 (7.5%) 195 (1.34%) 893 (6.16%)

Breast Cancer Risk and Exposure in Early Life to Polycyclic Aromatic Hydrocarbons Using Total Suspended Particulates as a Proxy Measure

Matthew R. Bonner, Daikwon Han, Jing Nie, Peter Rogerson,
John E. Vena, Paola Muti,Maurizio Trevisan,
Stephen B. Edge, and Jo L. Freudenheim

Cancer Epidemiol Biomarkers Prev 2005;14(1). January 2005

Polycyclic aromatic hydrocarbons (PAHs) are ubiquitous in the environment and present in air pollution. Early life exposure to PAHs may have particular importance in the etiology of breast cancer.

We conducted a population-based, case-control study of ambient PAH exposure in early life in relation to the risk of breast cancer. Total suspended particulates (TSP), a measure of ambient air pollution, was used as a proxy for PAH exposure.

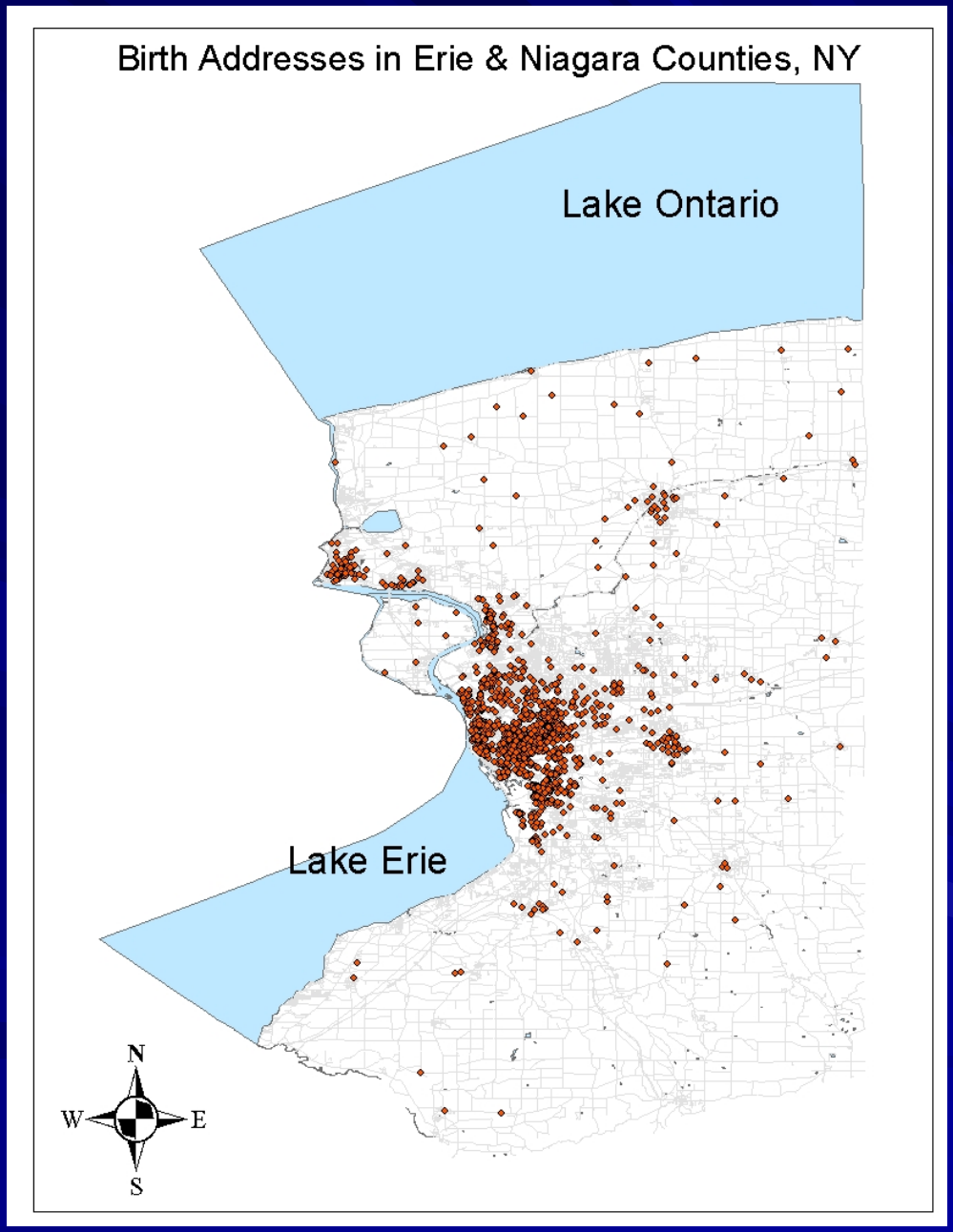
Cases were 1,166 women with histologically-confirmed, primary, incident breast cancer. Controls (*n*=2,105) were frequency matched by age, race, and county of residence to cases.

Annual average TSP concentrations (1959-1997) were obtained from the New York State Department of Environmental Conservation for Erie and Niagara Counties. Based on the monitor readings for each time period, prediction maps of TSP concentrations were generated with ArcGIS 8.0 (ESRI, Inc., Redlands, CA) using inverse distance squared weighted interpolation.

Statistical Analysis

Unconditional logistic regression was used to estimate odds ratios (OR) and 95% confidence intervals (95% CI). TSP concentrations were categorized into 4 levels (<84 ug/m3, 84-114 ug/m3, 115-140ug/m3, and >140ug/m3). The cut points for the categorical analyses were derived from the quartiles of the distribution of measurements of TSP concentrations in the 1960s.

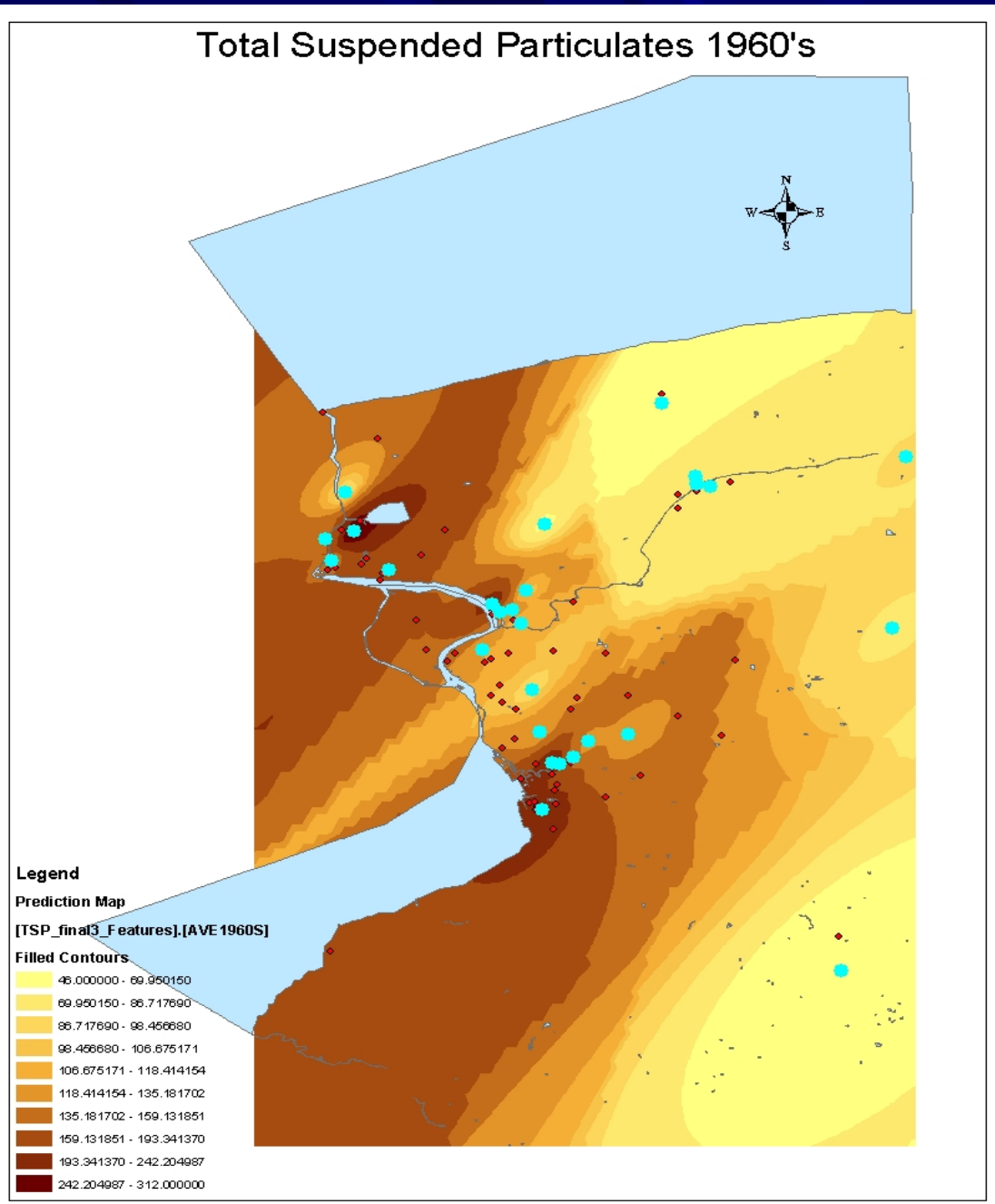
We also examined TSP concentrations on a continuous scale. Further, logistic quadratic spline regression with knots at 84 μg/m3 and 140 μg/m3 was used to graphically depict the exposure-response trend; the estimated probability of being a case was calculated from the quadratic spline regression equation. The values for the two knots in the spline regression were selected based on the previous categorical analysis. The end categories were restricted to linear segments to prevent instability.


Birth Addresses in Erie & Niagara Counties, NY
Lake Ontario
Lake Erie
N
W
E
S

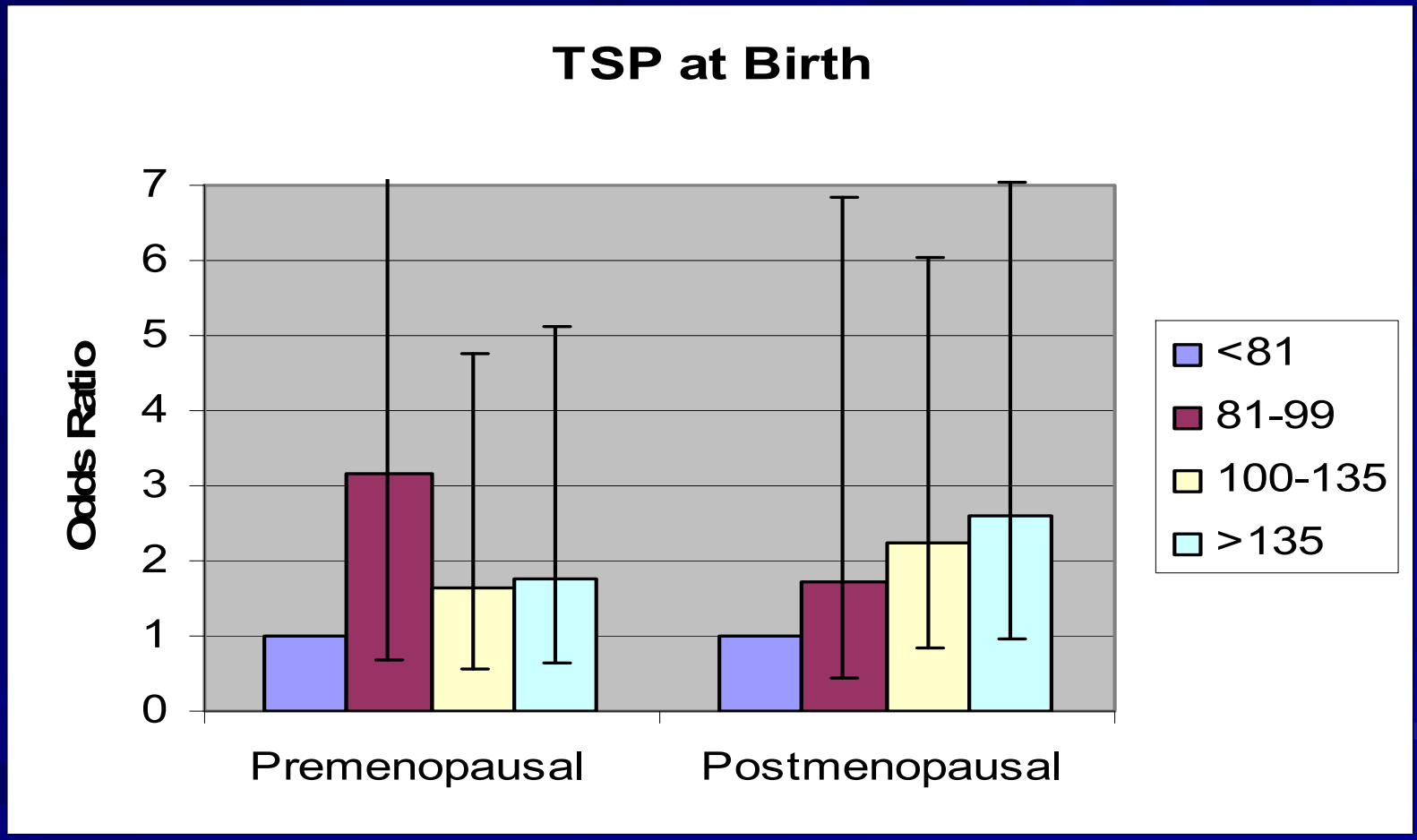
Study Question

- **Is exposure to high total suspended particulates associated with occurrence of breast cancer?**
  1. **Birth**
  2. **Menarche**
  3. **First birth**
  4. **20 years before diagnosis**
  5. **10 years before diagnosis**
  6. **Cumulative exposure**

Total Suspended Particulates 1960's



ORs and 95%CI: TSP at Birth Address



Adjusted for age, education, age at menarche, benign breast disease, parity, BMI, family hx. (Cases n=505; Controls n=804)

ORs and 95%CI: TSP at Birth Address- Postmenopausal women

TSP μg/m^3	Cases (n=345)	Controls (n=521)	Adjusted OR (95% CI)
<81	6	18	1.00
81-99	7	12	1.7 (0.4-6.8)
100-135	156	238	2.2 (0.8-6.0)
>135	176	253	2.6 (1.0-7.0)
P for trend			0.009

Adjusted for age, education, age at menarche, benign breast disease, parit

Figure 2. Estimated Probability of Being a Case for Postmenopausal Women by Total Suspended Particulate Concentration at Birth Address.

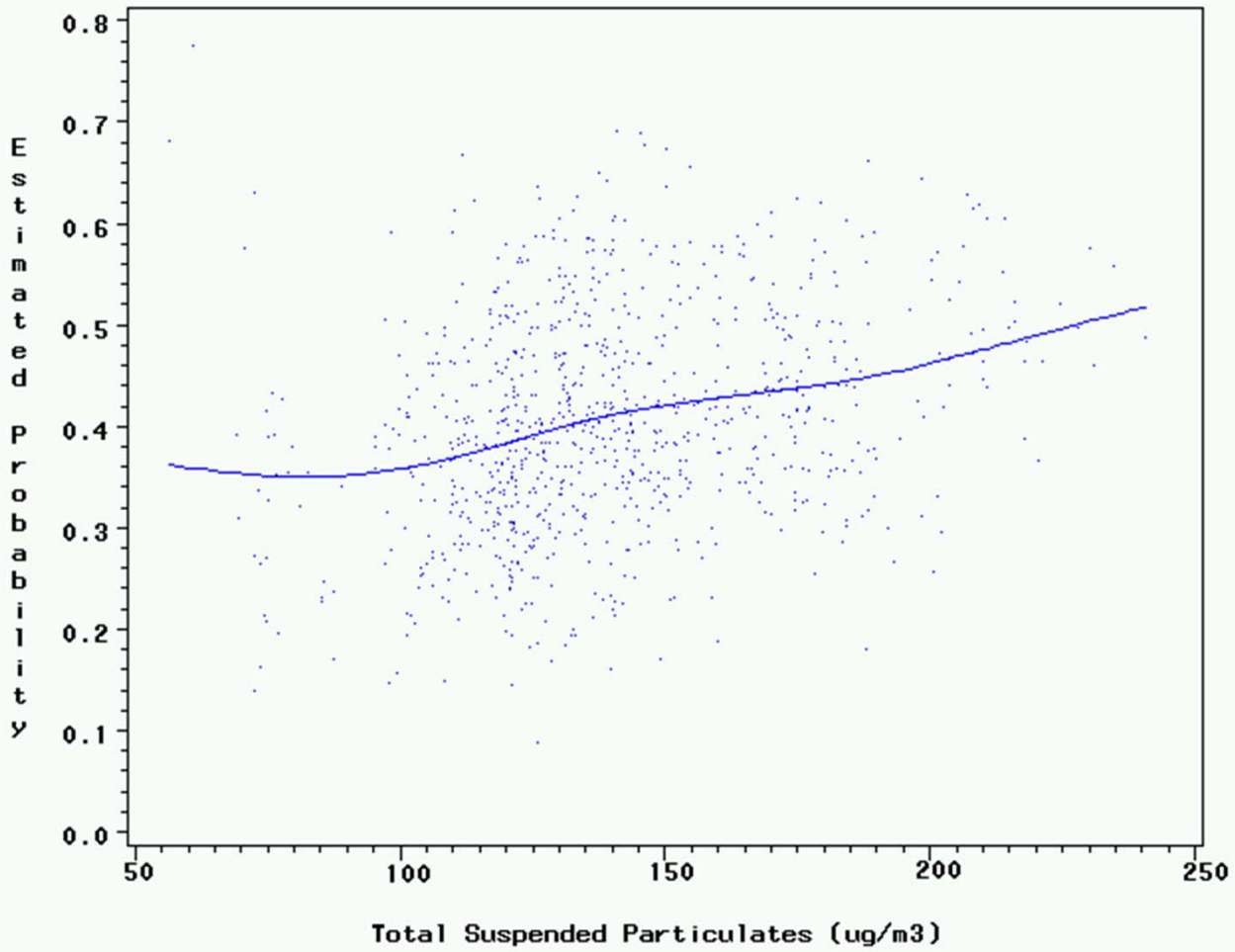
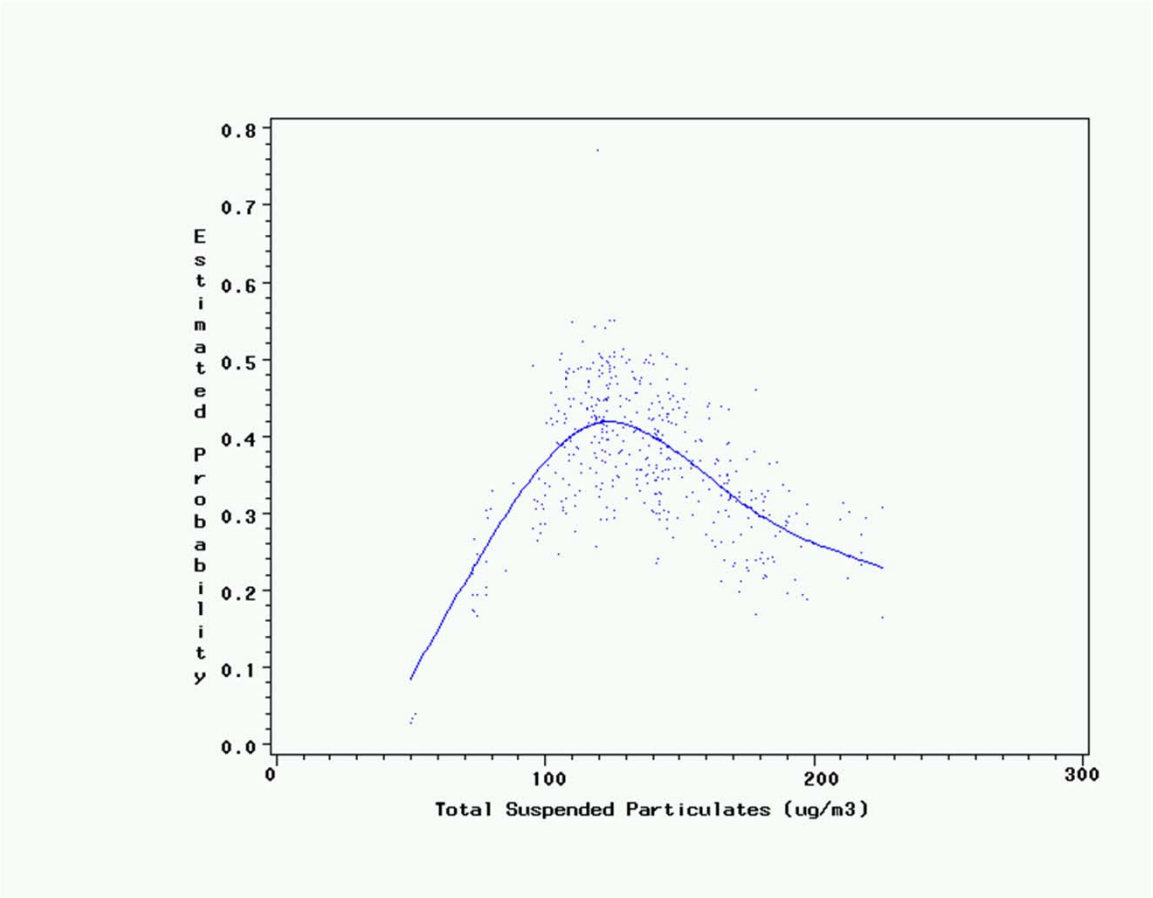


Figure 3. Estimated Probability of Being a Case for Premenopausal Women by Total Suspended Particulate Concentration at Birth Address.



Summary

- Exposure to high TSP concentrations appear to be associated with an increase in risk of breast cancer in postmenopausal women.
  - TSP at birth
- TSP exposure was not associated with premenopasual breast cancer.

International Journal of Health Geographics

Research

Assessing spatio-temporal variability of risk surfaces using residential history data in a case control study of breast cancer

Daikwon Han*[1,2], Peter A Rogerson[2,3], Matthew R Bonner[1], Jing Nie[1], John E Vena[4], Paola Muti[1], Maurizio Trevisan[1] and Jo L Freudenheim[1]

Address: [1]Department of Social and Preventive Medicine, University at Buffalo, Buffalo, NY 14214 USA, [2]Department of Geography and National Center for Geographic Information and Analysis, University at Buffalo, Buffalo, NY 14261 USA, [3]Department of Biostatistics, University at Buffalo, Buffalo, NY 14214 USA and [4]Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC 29208 USA

Email: Daikwon Han* - dhan@buffalo.edu; Peter A Rogerson - rogerson@buffalo.edu; Matthew R Bonner - bonnerm@mail.nih.gov; Jing Nie - jingnie@buffalo.edu; John E Vena - jvena@gwm.sc.edu; Paola Muti - muti@buffalo.edu; Maurizio Trevisan - trevisan@buffalo.edu; Jo L Freudenheim - jfreuden@buffalo.edu

* Corresponding author

Published: 12 April 2005

International Journal of Health Geographics 2005, **4**:9 doi:10.1186/1476-072X-4-9

This article is available from: http://www.ij-healthgeographics.com/content/4/1/9

Received: 28 February 2005
Accepted: 12 April 2005

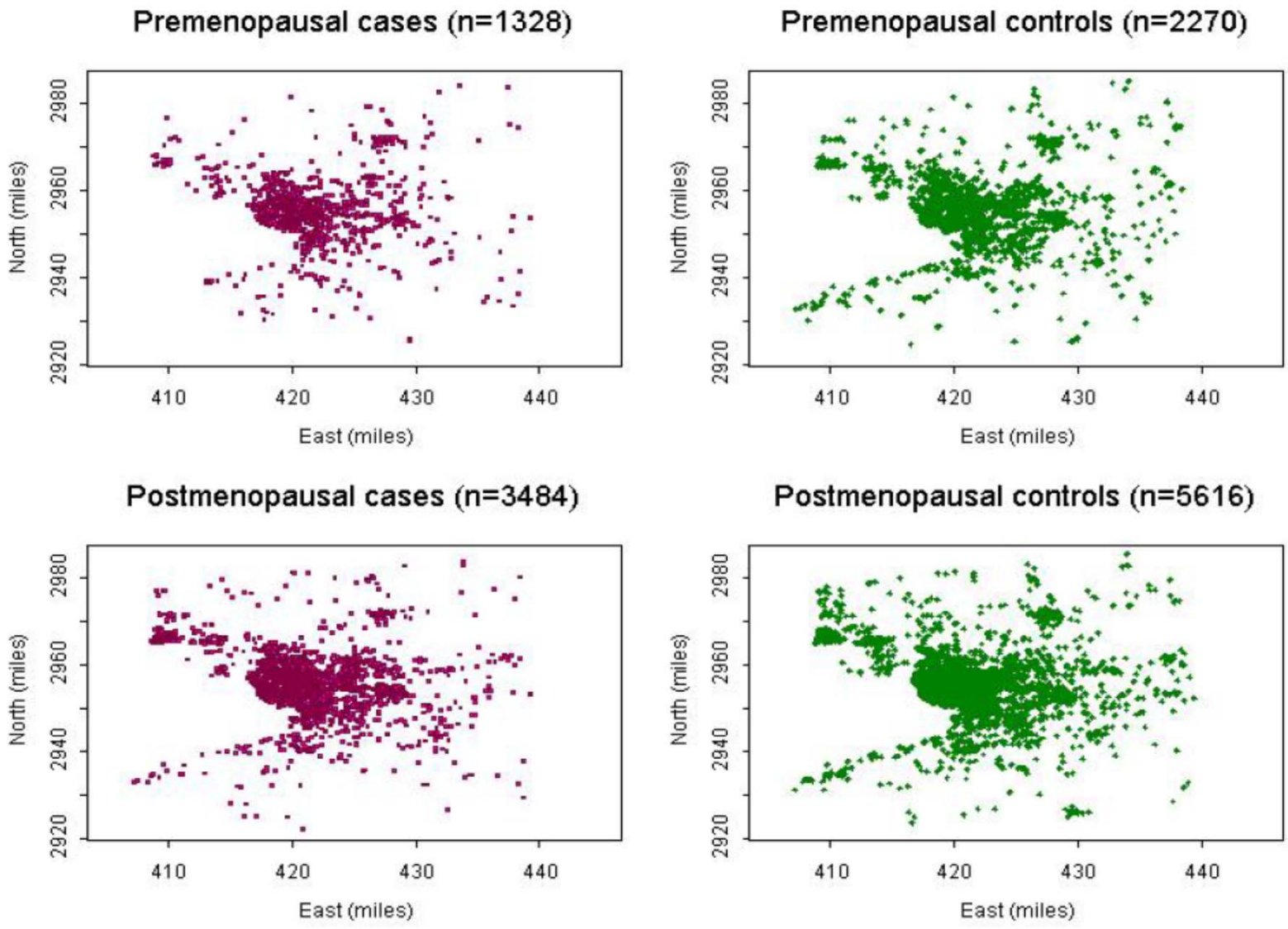


Figure 2

Geographic distribution of breast cancer in Western New York; Shown are all residential locations of breast cancer cases and controls by menopausal status included in the analysis. One dot indicates each residential location. The rectangular region was used as an approximate boundary of the study area instead of actual county boundary in Figure 1. East (x) and north (y) coordinates in projected Universal Traverse Mercator (UTM) miles.

Results: A GIS-based exploratory spatial analysis was applied, and spatio-temporal variability of those risk surfaces was evaluated using the standardized difference in density surfaces between cases and controls. The significance of the resulting risk surfaces was tested and reported as *p*-values. These surfaces were compared for premenopausal and postmenopausal women, and were obtained for each decade, from the 1940s to 1990s. We found strong evidence of clustering of lifetime residence for premenopausal women (for cases relative to controls), and a less strong suggestion of such clustering for postmenopausal women, and identified a substantial degree of temporal variability of the risk surfaces.

Conclusion: We were able to pinpoint geographic areas with higher risk through exploratory spatial analyses, and to assess temporal variability of the risk surfaces, thus providing a working hypothesis on breast cancer and environmental exposures. Geographic areas with higher case densities need further epidemiologic investigation for potential relationships between lifetime environmental exposures and breast cancer risk. Examination of lifetime residential history provided additional information on geographic areas associated with higher risk; limiting exploration of chronic disease clustering to current residence may neglect important relationships between location and disease.

Archives of Environmental & Occupational Health, Vol. 62, No. 1, 2007

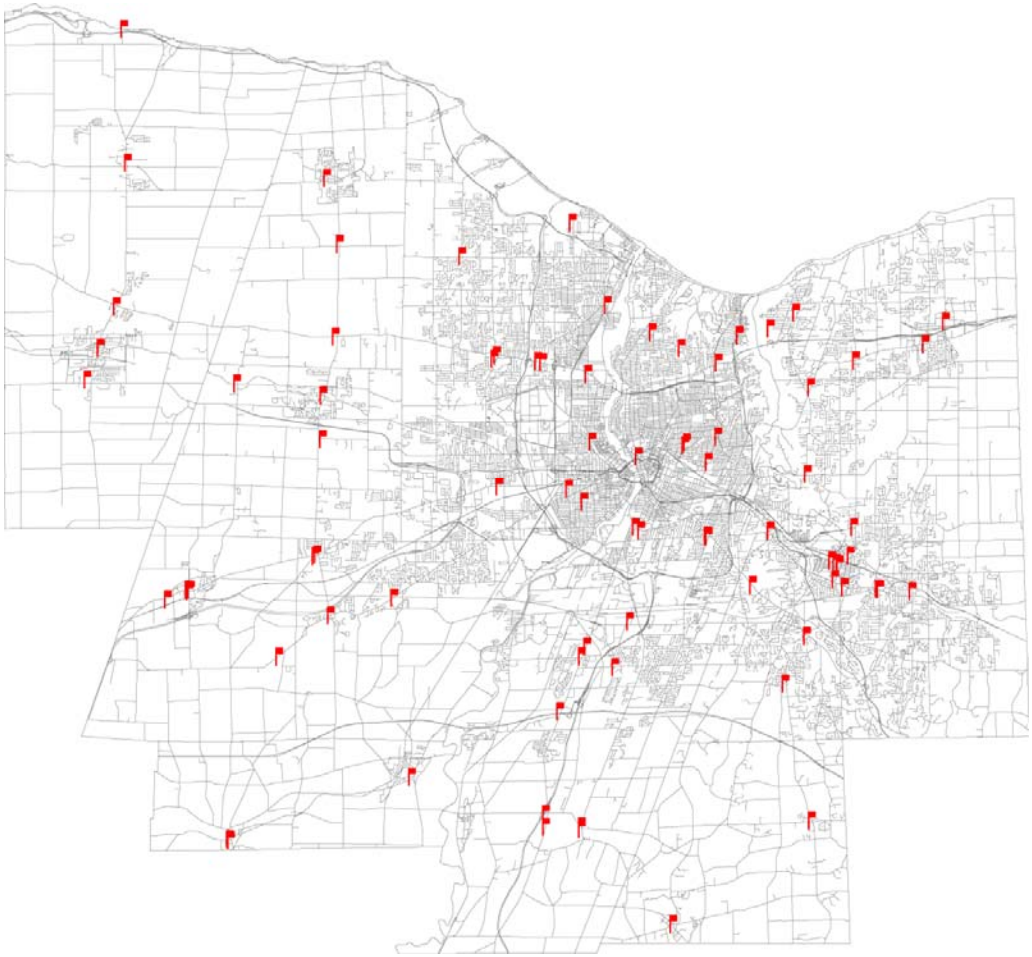
Case-Control Study of the Effects of Trihalomethanes on Urinary Bladder Cancer Risk

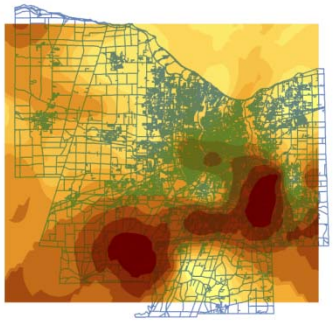
Gerald E. Bove Jr., PhD; Peter A. Rogerson, PhD; John E. Vena, PhD

ABSTRACT. In this research, the authors examined the relation between the estimated concentrations in drinking water of disinfectant byproduct (DBP) trihalomethanes (THMs) and the risk for urinary bladder cancer in a case-control study of 567 white men aged 35 to 90 years, in western New York State. They used logistic regression to estimate odds ratios and to assess the effects of THM consumption on cancer risk. Higher levels of consumption of THMs led to increased risk for cancer of the urinary bladder (total 551, a composite measure of THMs based upon method 551 developed by the US Envirnmental Protection Agency: OR = 2.34; 95% CI = 1.01–3.66). Results were most significant for bromoform (OR = 3.05; 95% CI = 1.51–5.69), and risk was highest (OR = 5.85; 95% CI = 1.93–17.46) for those who consumed the greatest amount of water at points within the distribution system with the oldest postdisinfected tap water.

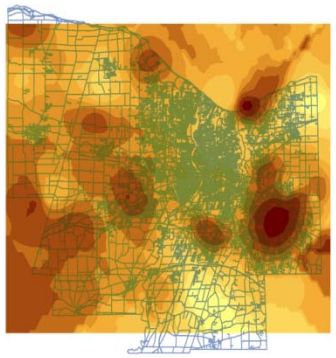
KEYWORDS: cancer risk, DBPs, drinking water, THMs, urinary bladder cancer

Locations of THM sample sites used in kriging

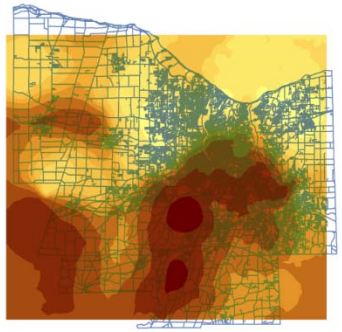




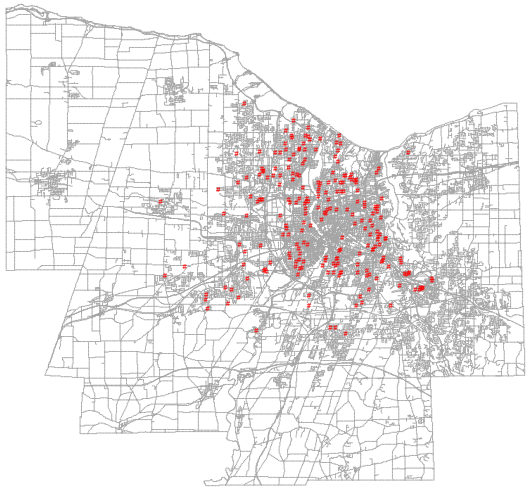
bromodichloromethane



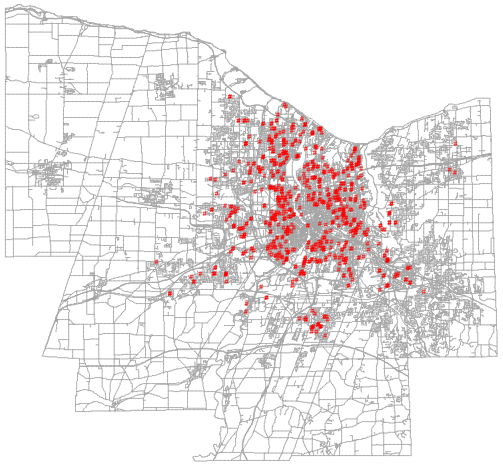
Chloroform



Total 551



Urinary bladder cases



Urinary bladder controls

Bladder Cancer

Higher levels of consumption of THMs led to increased risk for cancer of the urinary bladder (Total 551 (a composite measure of THMs) OR =2.34, 95 % CI = 1.01-3.66).

Results were most significant for Bromoform (OR = 3.05, 95 % CI = 1.51-5.69), and risk was highest (OR = 5.85, 95% CI = 1.93-17.46) for those who consumed the greatest amount of water at points within the distribution system with the oldest postdisinfection tap water.

International Journal of Health Geographics

Research

Case control study of the geographic variability of exposure to disinfectant byproducts and risk for rectal cancer

Gerald E Bove Jr[1], Peter A Rogerson*[1,2] and John E Vena[3]

Address: [1]Department of Geography, University at Buffalo, Wilkeson Hall, Buffalo, NY 14261, USA, [2]Department of Biostatistics, University at Buffalo, Farber Hall, Buffalo, NY 14261, USA and [3]Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, 800 Sumter Street, Columbia, SC 29208, USA

Email: Gerald E Bove - gebove@buffalo.edu; Peter A Rogerson* - rogerson@buffalo.edu; John E Vena - jvena@gwm.sc.edu

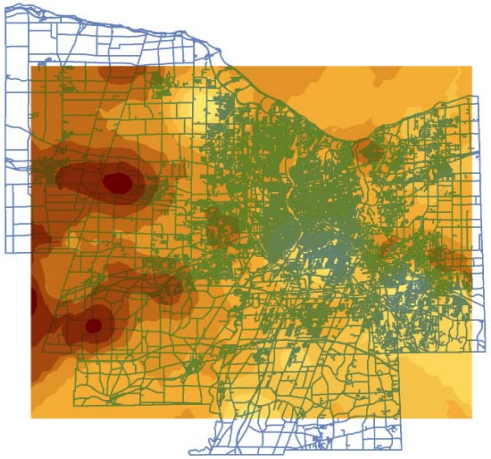
* Corresponding author

Published: 29 May 2007

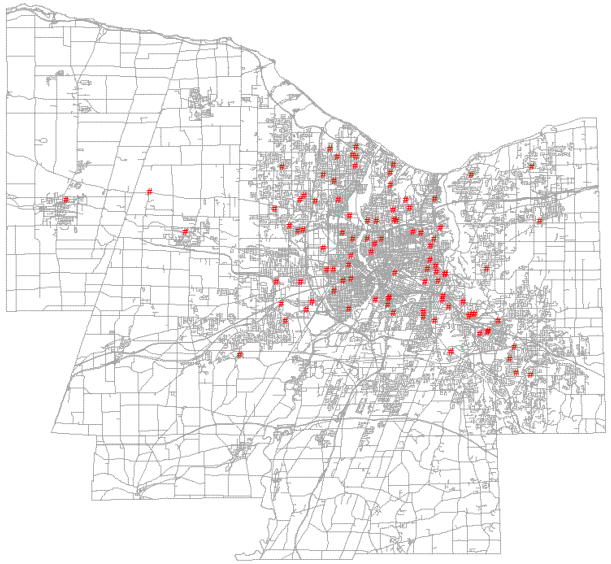
International Journal of Health Geographics 2007, **6**:18 doi:10.1186/1476-072X-6-18

Received: 15 February 2007
Accepted: 29 May 2007

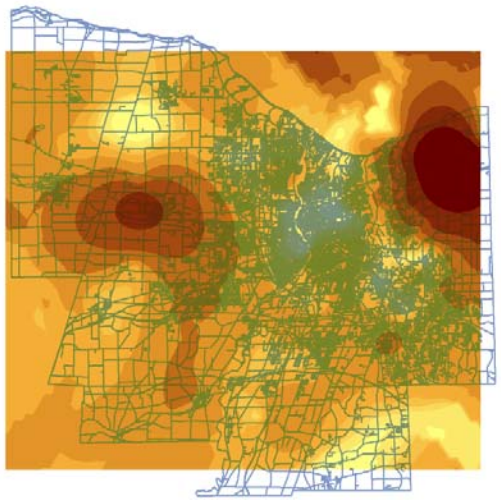
This article is available from: http://www.ij-healthgeographics.com/content/6/1/18



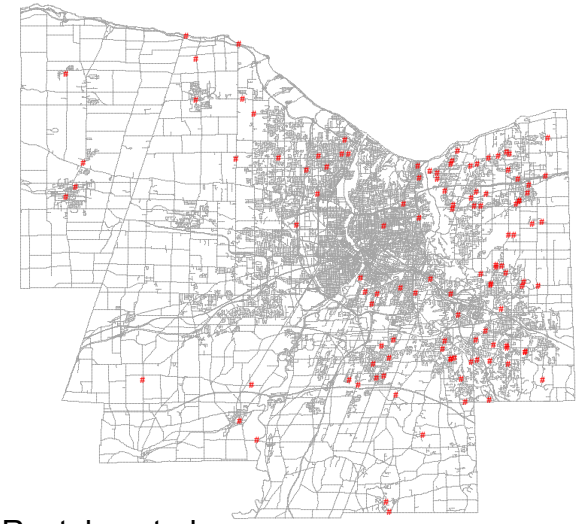
Chlorodibromomethane



Rectal cases



Bromoform



Rectal controls

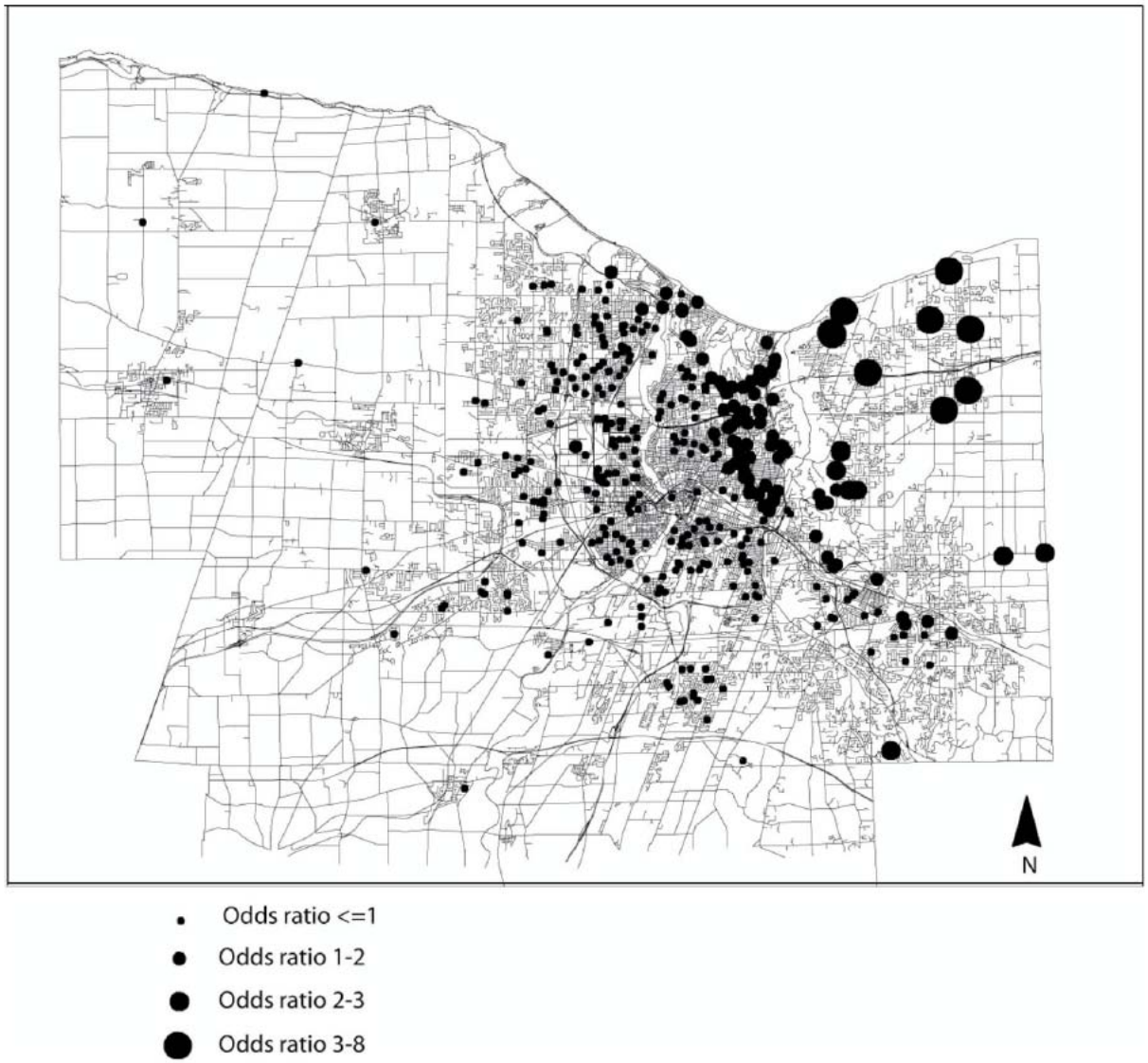
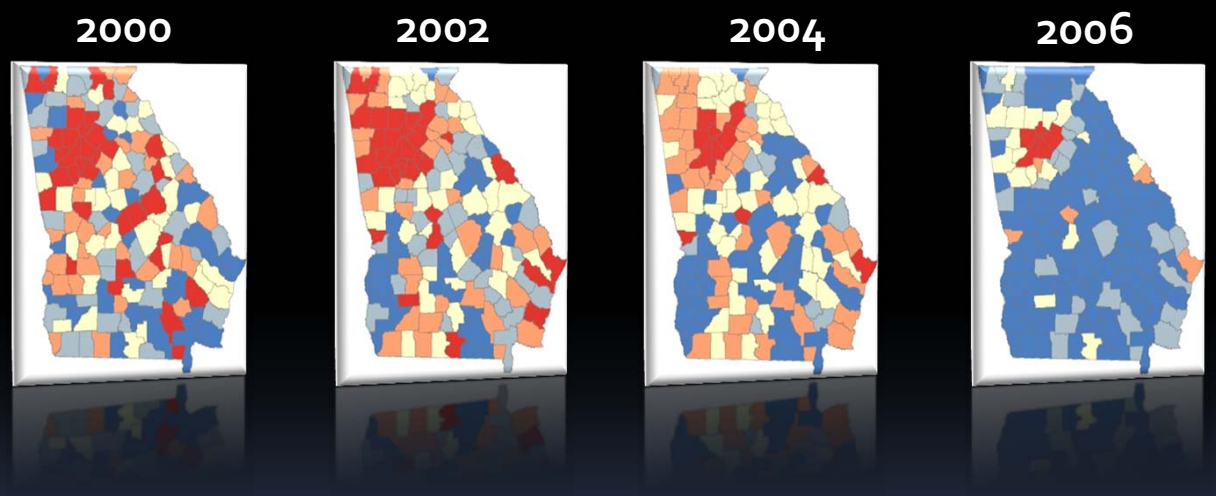


Figure 3
Individual odds ratios for rectal cancer risk for exposure to the THM bromoform. Note: Dependent variable determined as total daily ingestion of bromoform (ug/l) given as daily tap water intake (ug/l) and total bromoform contents of tap water (ug/l). Adjusted via assigning "average" values for covariates

- **Results**: Trihalomethane levels varied spatially within the county; increasing levels of the component bromoform (measured in ug/day) did correspond with an increase in odds ratios (OR = 1.85; 95% CI = 1.25 – 2.74) for rectal cancer.
- The highest quartiles of estimated consumption of bromoform (1.69–15.43 ug/day) led to increased risk for rectal cancer (OR = 2.32; 95% CI = 1.22–4.39).
- Two other THMs were marginally associated with an increase in risk – chlorodibromomethane (OR = 1.78, 95% CI = 1.00–3.19) and bromodichloromethane (OR = 1.15; 95% CI =1.00–1.32).

Conclusion: Levels of THMs in the water distribution system exhibited spatial variation that was partially due to variation in water age. We also observed a geographic pattern of increased risk of rectal cancer in areas with the highest levels of bromoform in the county.

Hierarchical Bayesian modeling of the spatio-temporal patterns of lung cancer incidence risk in Georgia, U.S. 2000-2007

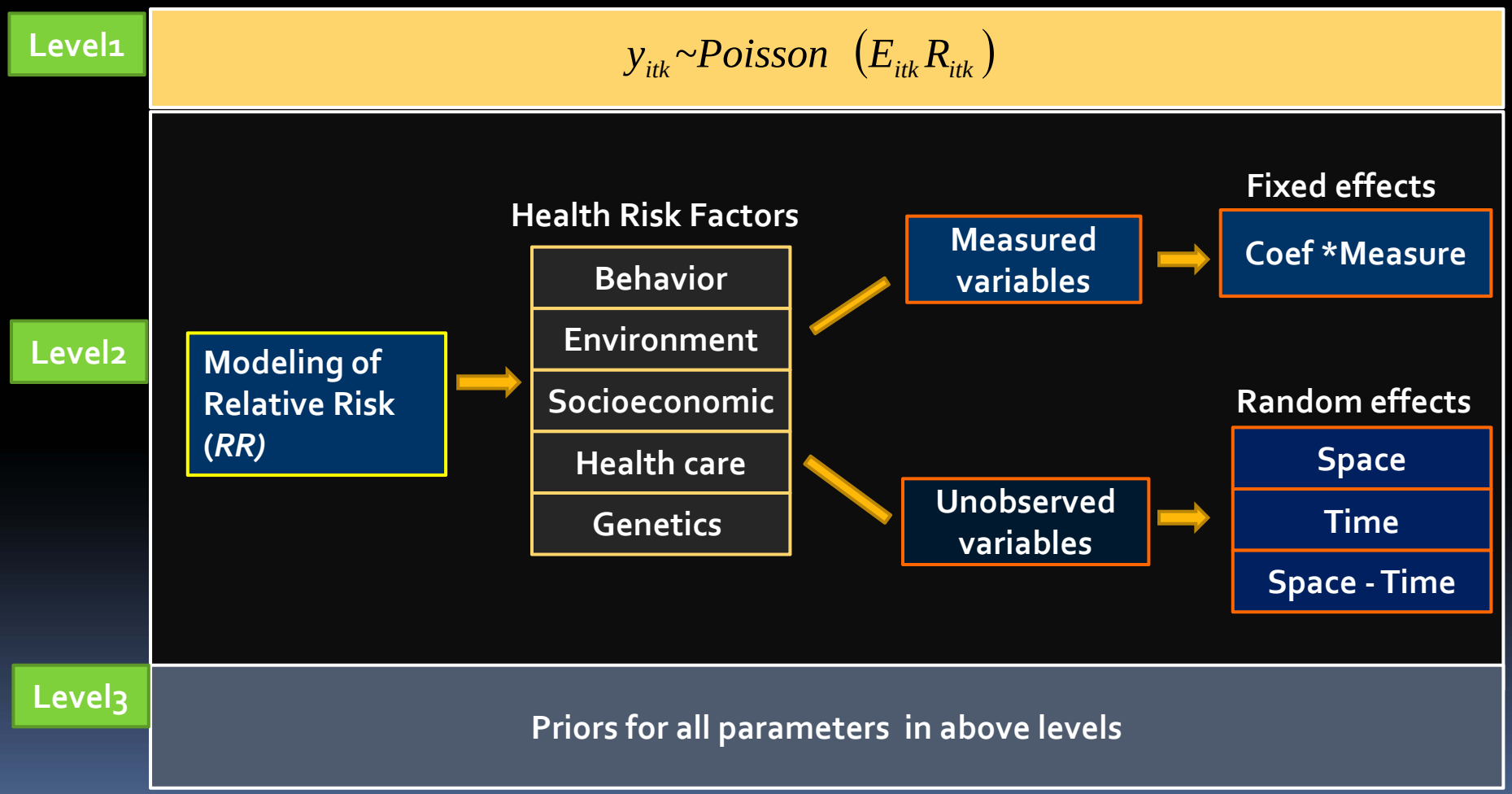


Ping Yin
Department of Geography
University of Georgia

- **Study Purposes:**
  - Obtain reliable spatio-temporal pattern of lung cancer incidence risk in Georgia at small scales
  - Understand the difference in the effects of socioeconomic status (SES) on the risk of each population subgroup

- **Study Design**:
  - **Time period**: *2000 -2007 (2-year analytical unit)*
  - **Population strata**:

    Race (white and black);
    Sex: male and female
  - **Data:** *Georgia Comprehensive Cancer Registry (GCCR) and census data*


Level1
$y_{itk} \sim Poisson\ (E_{itk} R_{itk})$
Level2
Modeling of Relative Risk (*RR*)
Health Risk Factors
Behavior
Environment
Socioeconomic
Health care
Genetics
Measured variables
Unobserved variables
Fixed effects
Coef *Measure
Random effects
Space
Time
Space - Time
Level3
Priors for all parameters in above levels

- **Model Selection**
  - Compare 5 joint models and 2 separate models
  - Deviance Information Criterion (DIC)

Autocorrelation is considered in the priors for all spatial and temporal random effects

- **Final Model : *Joint mapping***

Relative Risk at log scale | Intercept | SES effect | Shared spatial random effects | Specific spatial random effects | Specific spatio-temporal interaction

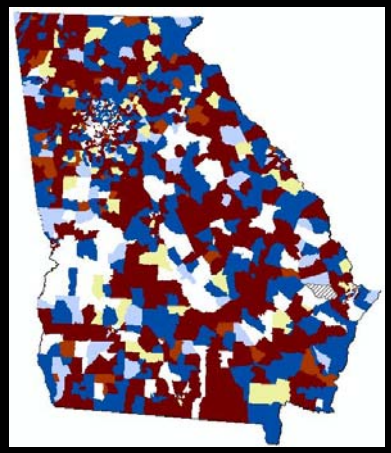
$$log\,(R_{itk}) = \alpha_k + \boldsymbol{\beta}_k^T \boldsymbol{x}_i + \delta_{1,k}\phi_i + \delta_{2,k}\varsigma_t + \lambda_{ik} + \xi_{tk} + \omega_{itk}$$

Shared temporal random effects | Specific temporal random effects

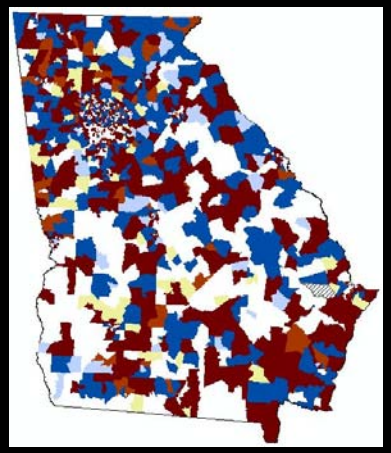
Tract index: *i=1, 2, …1618;*
Time period index: t=1, 2, 3, 4;
Population group index: k=1, 2, 3, 4;

Standardized Incidence Rates (SIR)2000-2001

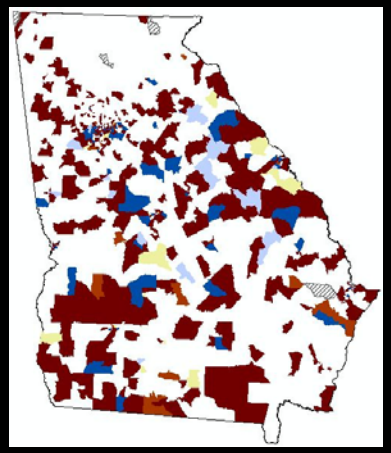
SIR= # Observed cases/ # Expected cases



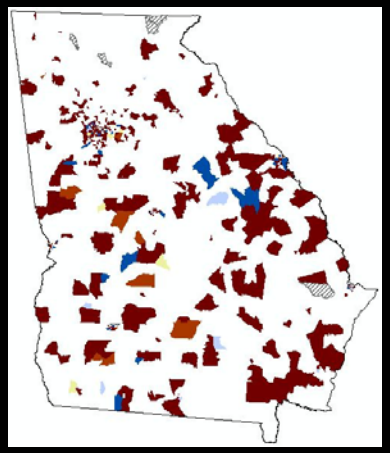
White Male



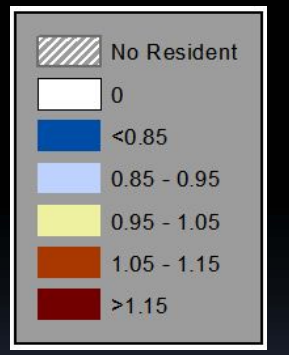
White Female

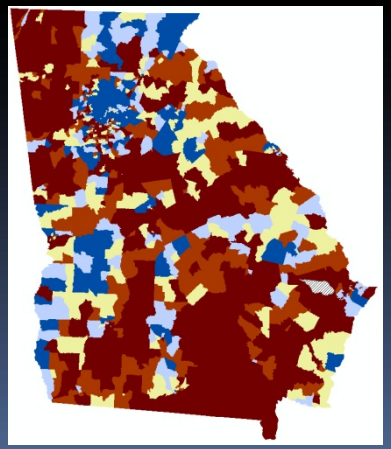


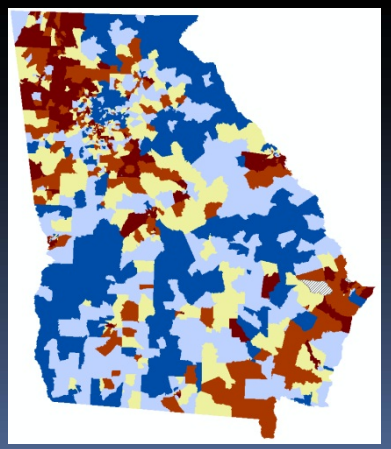
Black Male

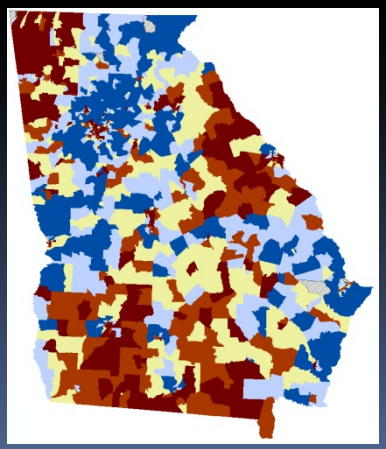


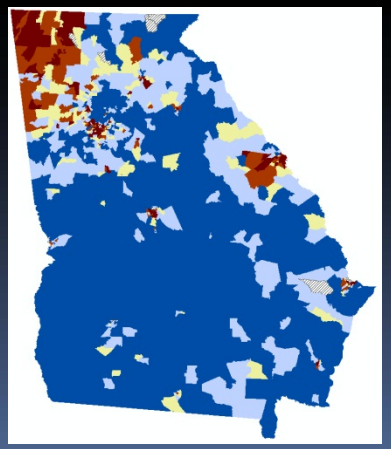
Black Female





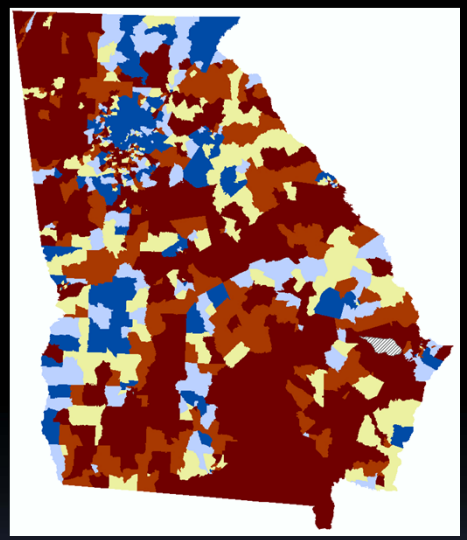




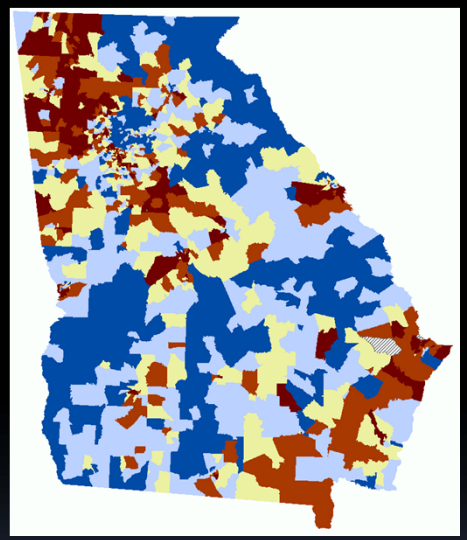


Relative Risks (RR) from Modeling 2000-2001

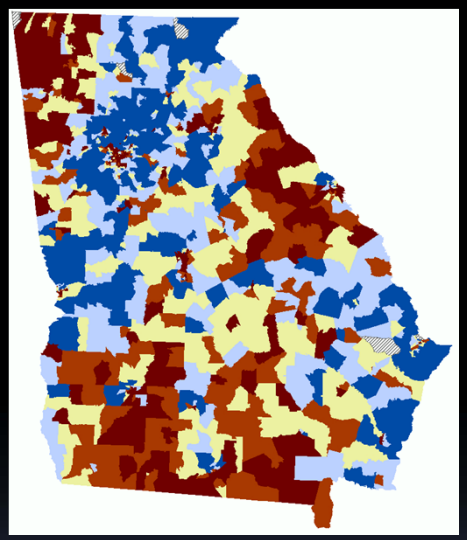
Relative Risk in 2000-2001



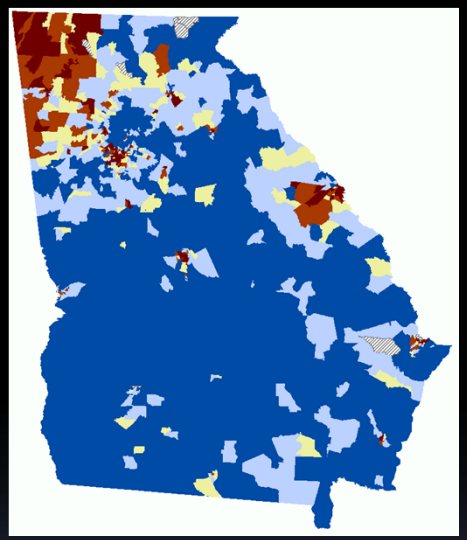
White Male



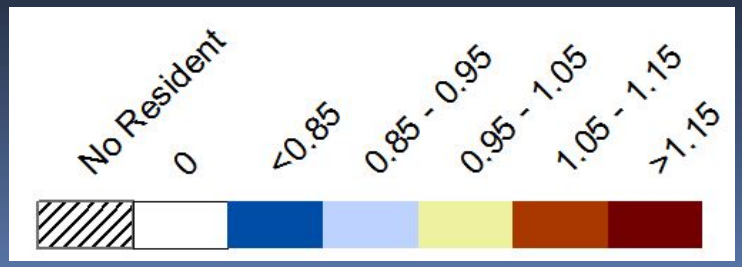
White Female



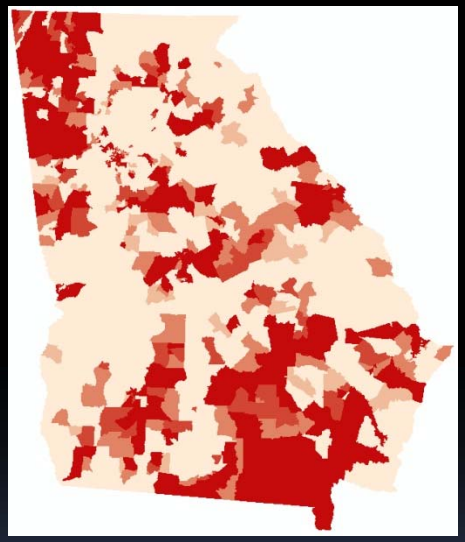
Black Male



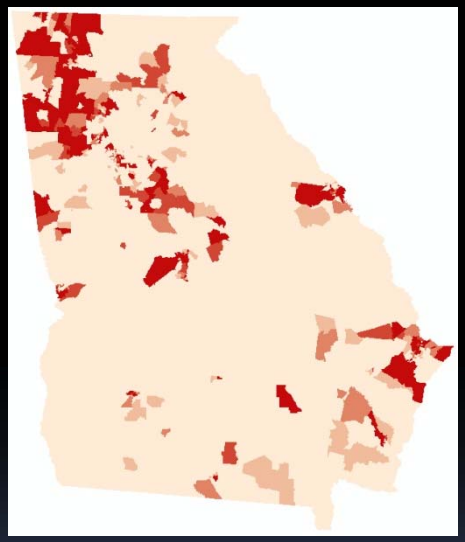
Black Female



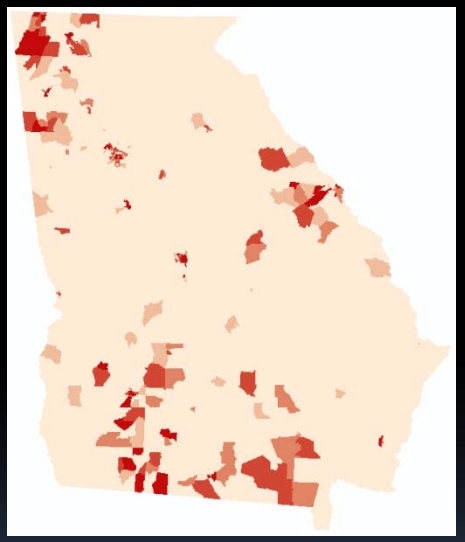
Frequency of Elevated Relative Risks for Each Tract during 2000-2007



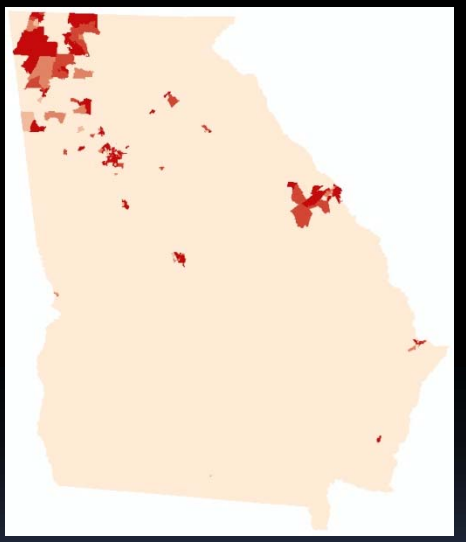
White Male



White Female

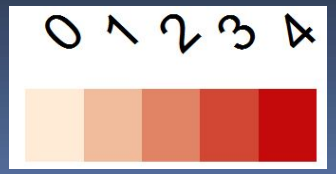


Black Male



Black Female

Elevated Risk Criterion:
Prob(RR>1) > 0.8



- Bayesian modeling with shared components makes smooth risk maps by borrowing strengths from neighboring regions and time periods, as well as other population groups.
- Northwest Georgia has stably high elevated lung cancer incidence relative risk for all races/sexes over 2000-2007.
- The SES effect on lung cancer incidence relative risk has a larger gradient in males, especially white males.

- Smoking data (if available) are expected to improve the modeling.
- SES is assumed static in this research and the long latency of cancers makes the SES 20 years prior to the death may be important.
- Estimation of population at risk could cause errors in the modeling.

Wagner SE, Hurley DB, Hebert JR, McNamara C, Bayakly AR, Vena JE. Cancer mortality-to-incidence ratios in Georgia: describing racial cancer disparities and potential geographical determinants *Cancer*

CANCER MORTALITY-TO-INCIDENCE RATIOS IN GEORGIA

The Mortality-to-Incidence Ratio

- Mortality-to-incidence ratio (MIR)

$$\text{MIR} = \frac{\text{(age-adjusted mortality cancer rate)}}{\text{(age-adjusted incidence cancer rate)}}$$

- AKA
  - Fatality ratio
  - Fatality, given incidence
  - 1/survival, given incidence

Introduction

- Racial disparities in cancer outcomes are large
  - Mortality
    - US: Blacks 25% higher mortality than other (all cancers)
    - Disparities are very large
    - Blacks have more aggressive tumors
  - Incidence
    - Usually higher in Blacks, but not always
      - Breast cancer
    - Disparities are present, but not as large

Methods

- Cancer incidence and mortality data by health district, 2003-2007
  - All sites combined, lung & bronchus, colorectal, female breast, oral, cervical, prostate
- Population data: NCHS bridged population estimates
- MIRs and 95% CI's generated by site, race, sex
- MIRs were mapped & compared to geographic health factors (County Health Rankings)

Health Factors

- Health Behaviors
  - Tobacco use; diet & exercise; alcohol use; unsafe sex
- Clinical Care
  - Access to care; quality of care
- Socioeconomic Factors
  - Education; employment; income; family & social support; community safety
- Physical Environment
  - Environmental quality; built environment

Health Factor Analysis

- County-level Z-scores averaged by health district
  - Positive Z-score: “greater risk” for worse health outcomes than average GA counties
  - Negative Z-score: “lower risk” for worse health outcomes than average GA counties
- Mapped by health district
- Correlation analysis between Z-scores and MIRs by health district

Results

- 186,419 incidence cancers (all sites)
- 71,533 cancer deaths (all sites)

Table 1. Georgia Mortality-to-Incidence Ratios and 95% Confidence Intervals for All Cancer Sites Combined and For Specified Cancer Sites by Race and Sex, 2003-2007

	Blacks		Whites		
Cancer Site: Sex Subgroup	MIR	95% CI	MIR	95% CI	Black:White Ratio[a]
All sites combined					
Overall[b]	0.450	0.442, 0.459	0.401	0.396, 0.405	1.122
Women[b]	0.443	0.432, 0.455	0.373	0.368, 0.379	1.188
Men[b]	0.471	0.458, 0.483	0.432	0.426, 0.439	1.090
Ratio of men to women	1.063	—	1.158	—	—
Female breast					
Women[b]	0.255	0.242, 0.269	0.181	0.175, 0.188	1.408
Cervix					
Women[b]	0.423	0.362, 0.492	0.279	0.247, 0.315	1.516
Colon and rectum					
Overall[b]	0.407	0.386, 0.430	0.343	0.331, 0.355	1.187
Women[b]	0.397	0.369, 0.427	0.339	0.322, 0.356	1.171
Men[b]	0.426	0.391, 0.463	0.352	0.335, 0.370	1.210
Ratio of men to women	1.073	—	1.038	—	—
Lung and bronchus					
Overall	0.793	0.761, 0.826	0.770	0.754, 0.785	1.029
Women	0.749	0.701, 0.799	0.721	0.699, 0.744	1.038
Men	0.827	0.783, 0.873	0.812	0.790, 0.834	1.018
Ratio of men to women	1.104	—	1.126	—	—
Oral cavity					
Overall[b]	0.333	0.291, 0.380	0.213	0.197, 0.230	1.563
Women	0.225	0.172, 0.292	0.215	0.187, 0.248	1.047
Men[b]	0.398	0.337, 0.469	0.220	0.199, 0.243	1.809
Ratio of men to women	1.769	—	1.023	—	—
Prostate					
Men[b]	0.242	0.228, 0.256	0.156	0.149, 0.163	1.551

Abbreviations: CI, confidence interval; MIR, mortality-to-incidence ratio.

[a] The black:white ratio is the MIR for blacks divided by the MIR for whites.

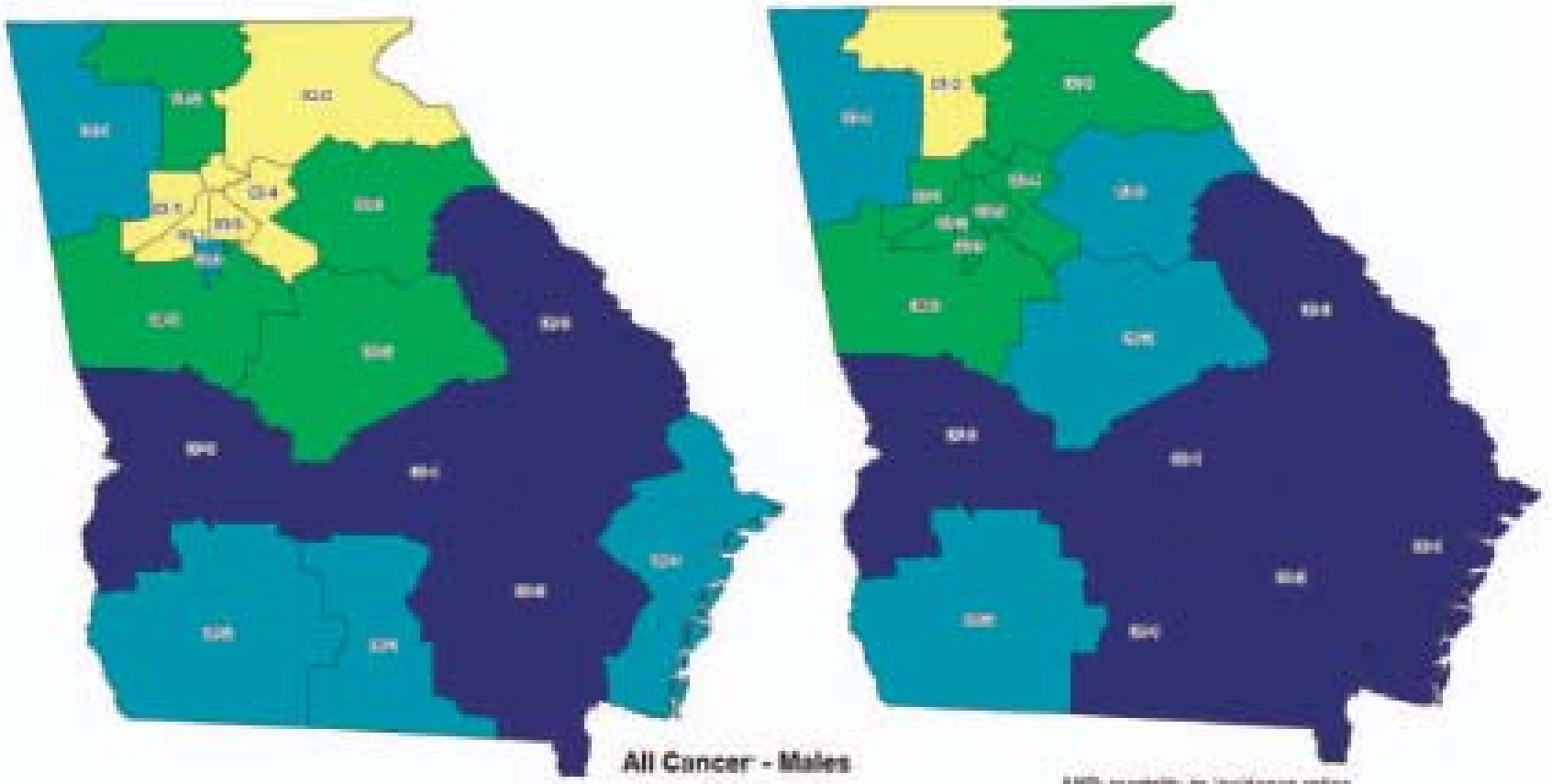
Mapping Display Cutoff Points

- Category 1: mean MIR for whites nationally
- Category 2: upper bound 10% higher than upper bound of Category 1
- Category 3: upper bound 20% higher than upper bound of Category 1
- Category 4: upper bound >20% higher than upper bound of Category 1

A

Age-Adjusted* All Cancer Sites Mortality-to-Incidence Ratios for White Males Compared to National Mean MIR**, 2003-2007

Age-Adjusted* All Cancer Sites Mortality-to-Incidence Ratios for Black Males Compared to National Mean MIR**, 2003-2007



All Cancer - Males

Categories for displaying MIR**

- Category 1 (<=0.410000)
- Category 2 (0.410001 - 0.450000)
- Category 3 (0.450001 - 0.490000)
- Category 4 (>0.490001)
- Georgia Public Health District Boundaries

MIR: mortality-to-incidence ratios.

*Rates are per 100,000 population and are age-adjusted to the 2000 US standard population.

**The upper bound of Category 1 is the mean MIR for all cancer sites combined among males for Whites nationally.

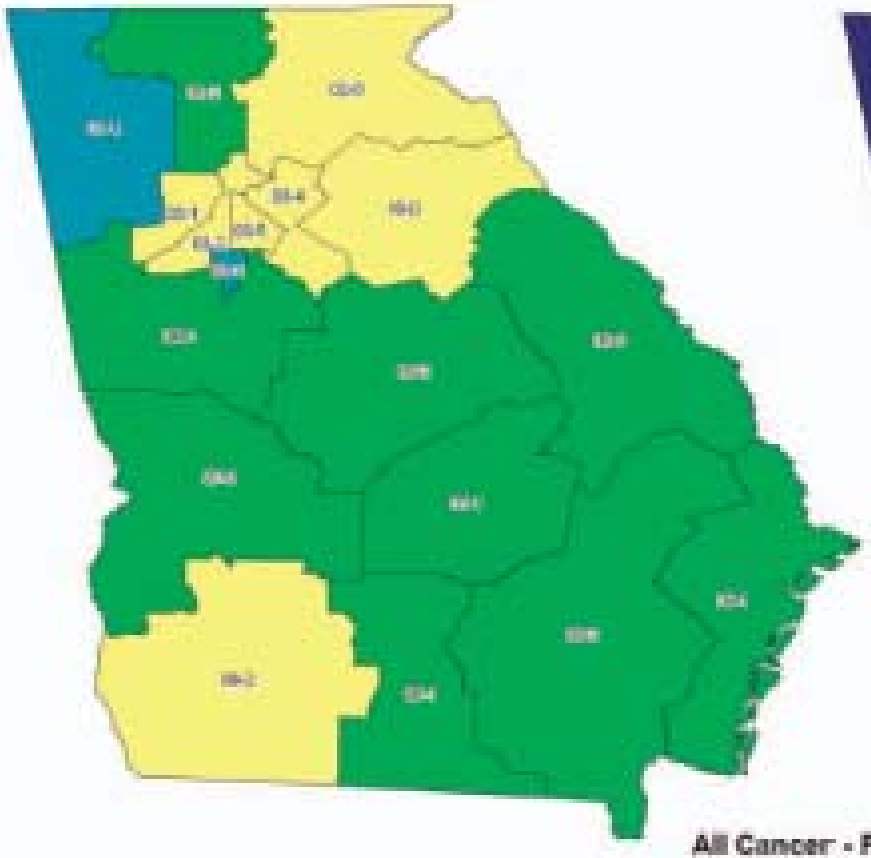
Data Sources:
Cancer incidence data: GA Comprehensive Cancer Registry
Cancer mortality data: GA Vital Statistics Records
Cut Points: US Cancer Statistics Working Group
Population data: North American Association of Cancer Registries

0 35 70 140 Miles

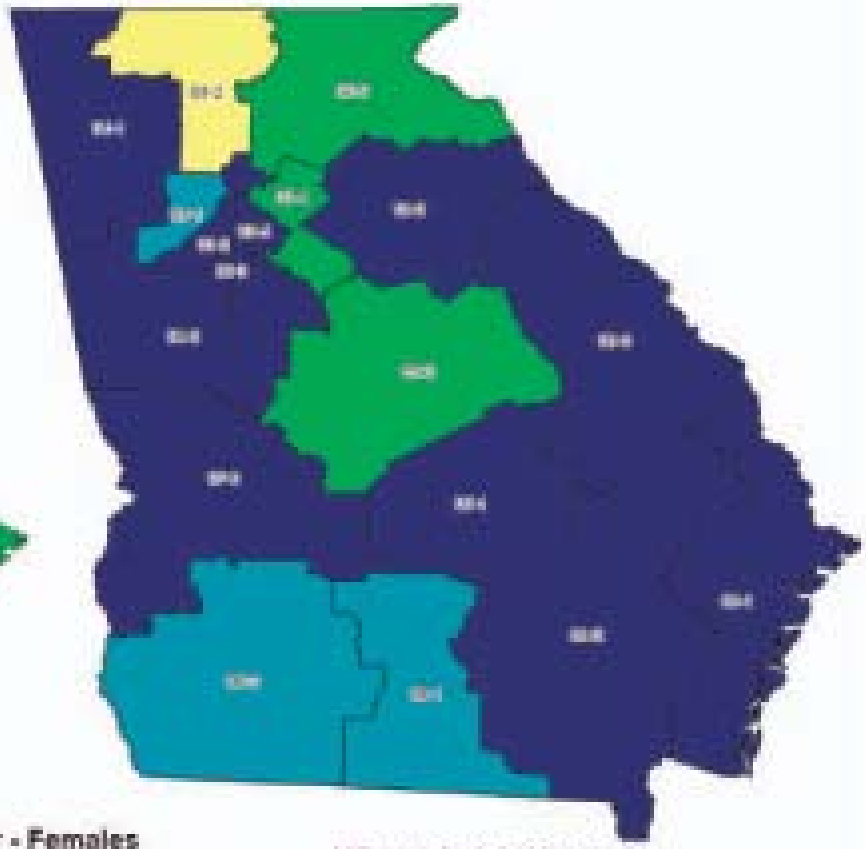
N

B

Age-Adjusted* All Cancer Sites Mortality-to-Incidence Ratios for White Females Compared to National Mean MIR**, 2003-2007



Age-Adjusted* All Cancer Sites Mortality-to-Incidence Ratios for Black Females Compared to National Mean MIR**, 2003-2007



Data Sources:
Cancer incidence data: GA Comprehensive Cancer Registry
Cancer mortality data: GA Vital Statistics Records
Cut Points: US Cancer Statistics Working Group
Population data: North American Association of Cancer Registries

All Cancer - Females

Categories for displaying MIR**

- Category 1 (=<0.370000)
- Category 2 (0.370001 - 0.410000)
- Category 3 (0.410001 - 0.440000)
- Category 4 (>0.440000)
- Georgia Public Health District Boundaries

MIR: mortality-to-incidence ratios.

*Rates are per 100,000 population and are age-adjusted to the 2000 US standard population.

**The upper bound of Category 1 is the mean MIR for all cancer sites combined among females for Whites nationally.

0 35 70 140 Miles

N

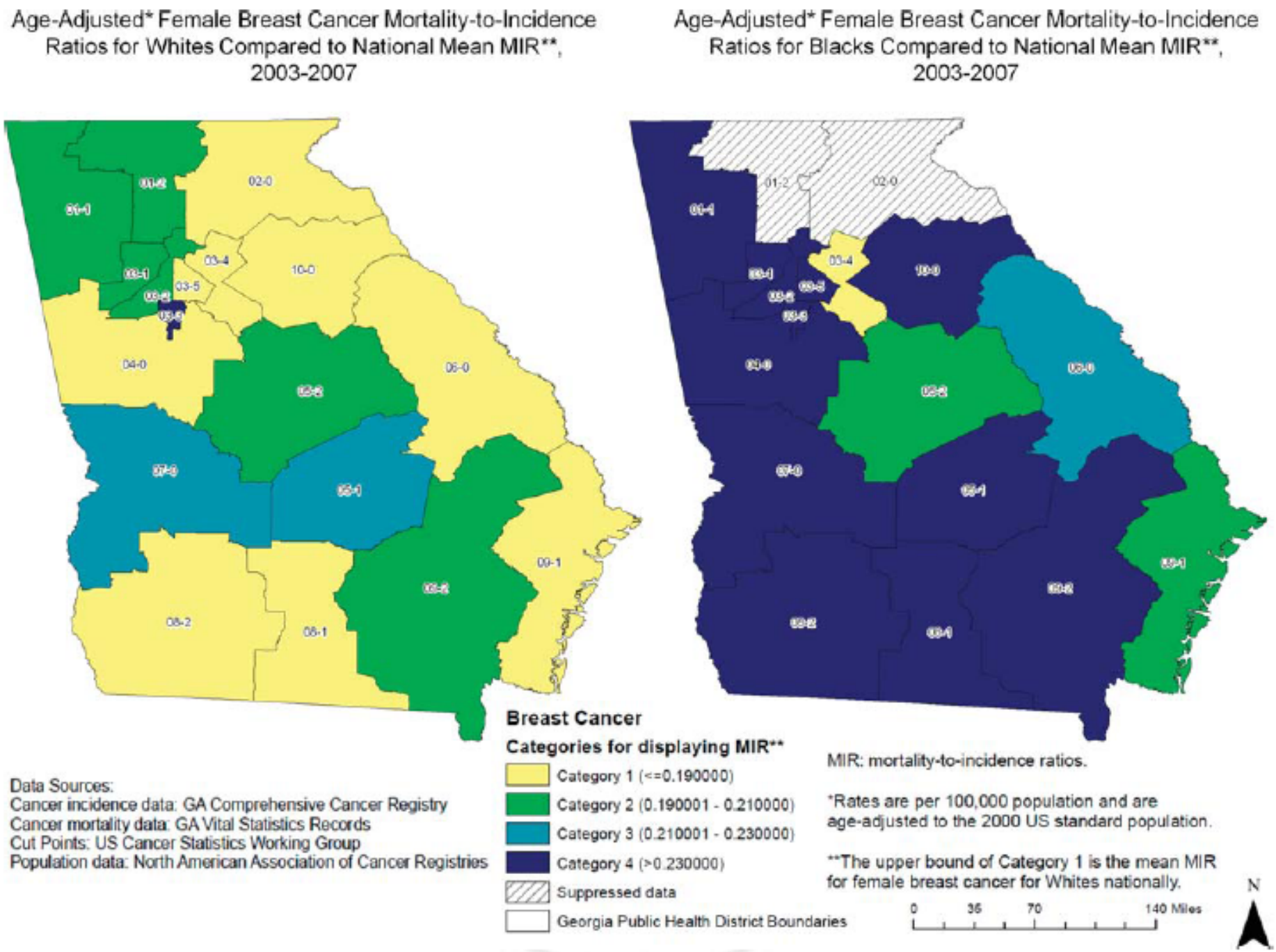


Figure 2. Mortality-to-incidence ratios are illustrated by Georgia public health district for female breast cancer.

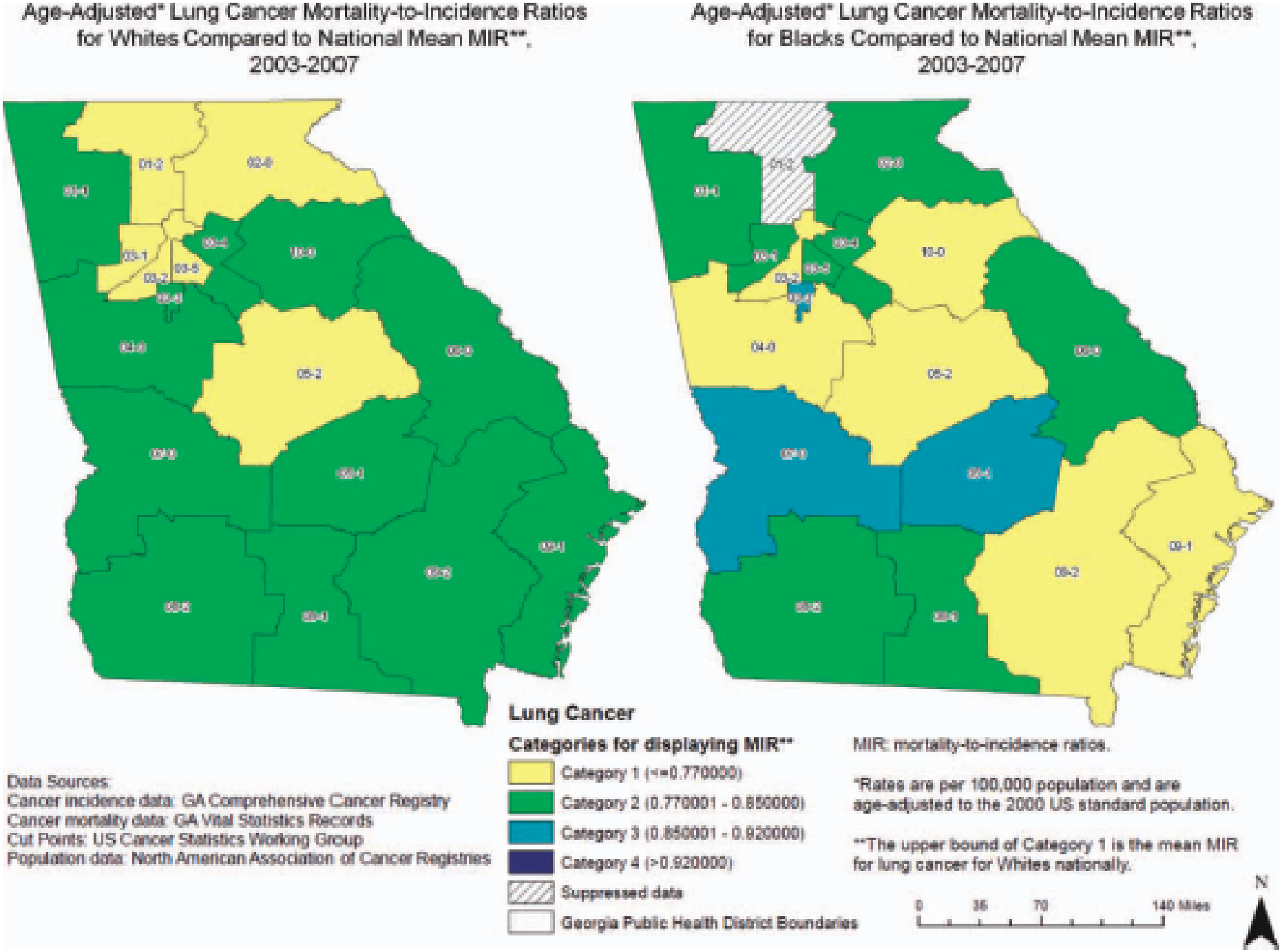


Figure 5. Mortality-to-incidence ratios are illustrated by Georgia public health district for lung cancer.

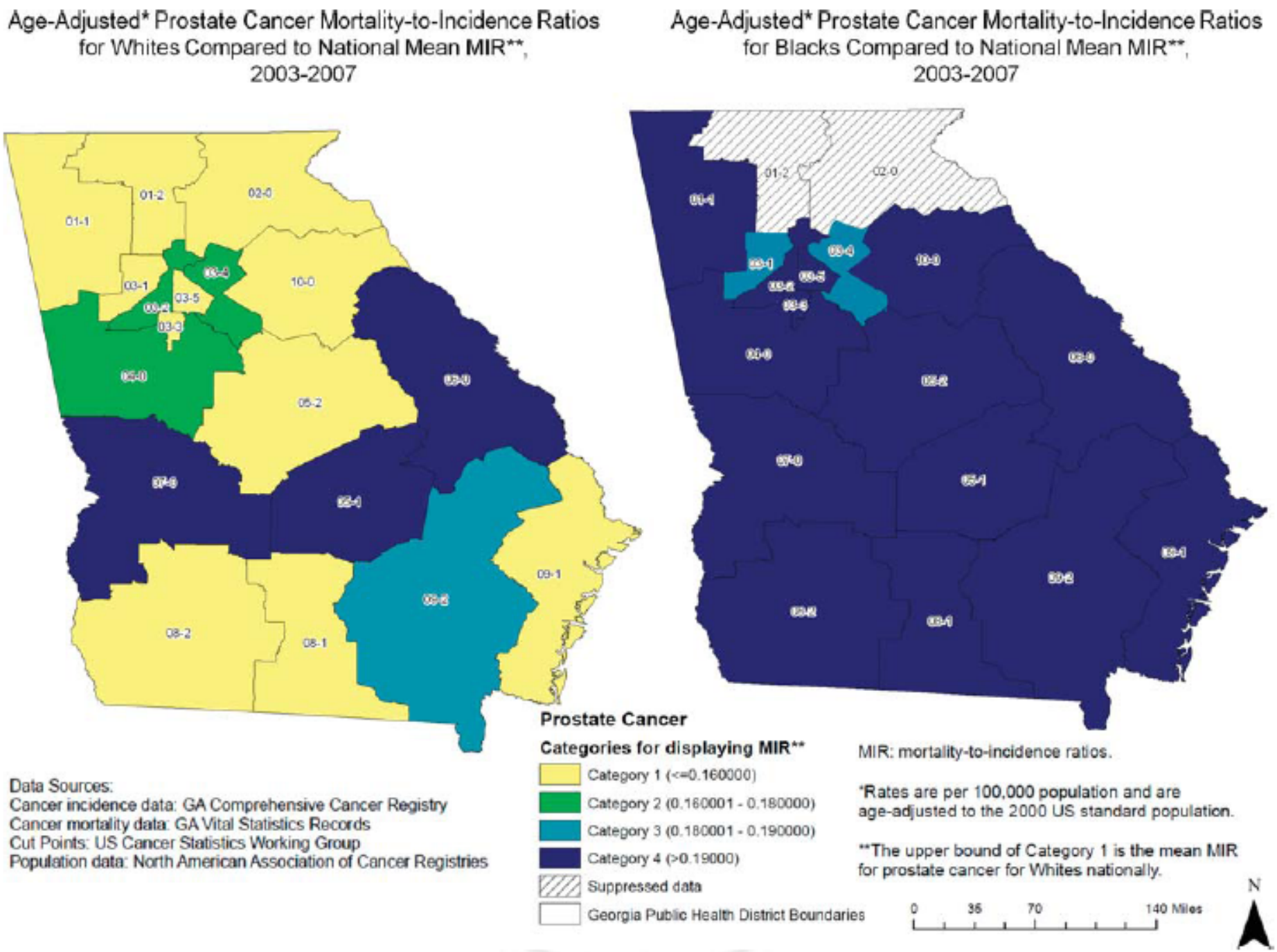


Figure 7. Mortality-to-incidence ratios are illustrated by Georgia public health district for prostate cancer.

Table 2. Correlation Coefficients Between Health District Health Factor Groupings and Average Health District Mortality-to-Incidence Ratios for All Cancer Sites Combined and for Specified Cancer Sites by Race and Sex, 2003-2007

	Blacks			Whites		
Cancer Site: Sex Subgroup	No.	ρ	*P*[a]	No.	ρ	*P*
Health behavior						
All sites combined						
Overall	18	0.463	0.03	18	0.860	<0.0001
Women	18	0.287	0.13	18	0.713	<.0001
Men	18	0.549	0.01	18	0.864	<.0001
Female breast						
Women	16	0.010	0.49	18	0.151	0.28
Cervix						
Women	5	0.443	0.23	7	−0.058	0.55
Colon and rectum						
Overall	17	−0.224	0.81	18	0.311	0.11
Women	15	−0.398	0.93	18	0.082	0.38
Men	14	0.060	0.42	18	0.311	0.11
Lung and bronchus						
Overall	17	0.394	0.06	18	0.708	<.0001
Women	16	−0.282	0.86	18	0.259	0.15
Men	17	−0.156	0.73	18	0.805	<.0001
Oral cavity						
Overall	7	0.423	0.17	16	0.326	0.11
Women	0	—	—	5	0.074	0.46
Men	4	0.709	0.15	14	0.198	0.25
Prostate						
Men	16	0.477	0.03	18	0.251	0.16

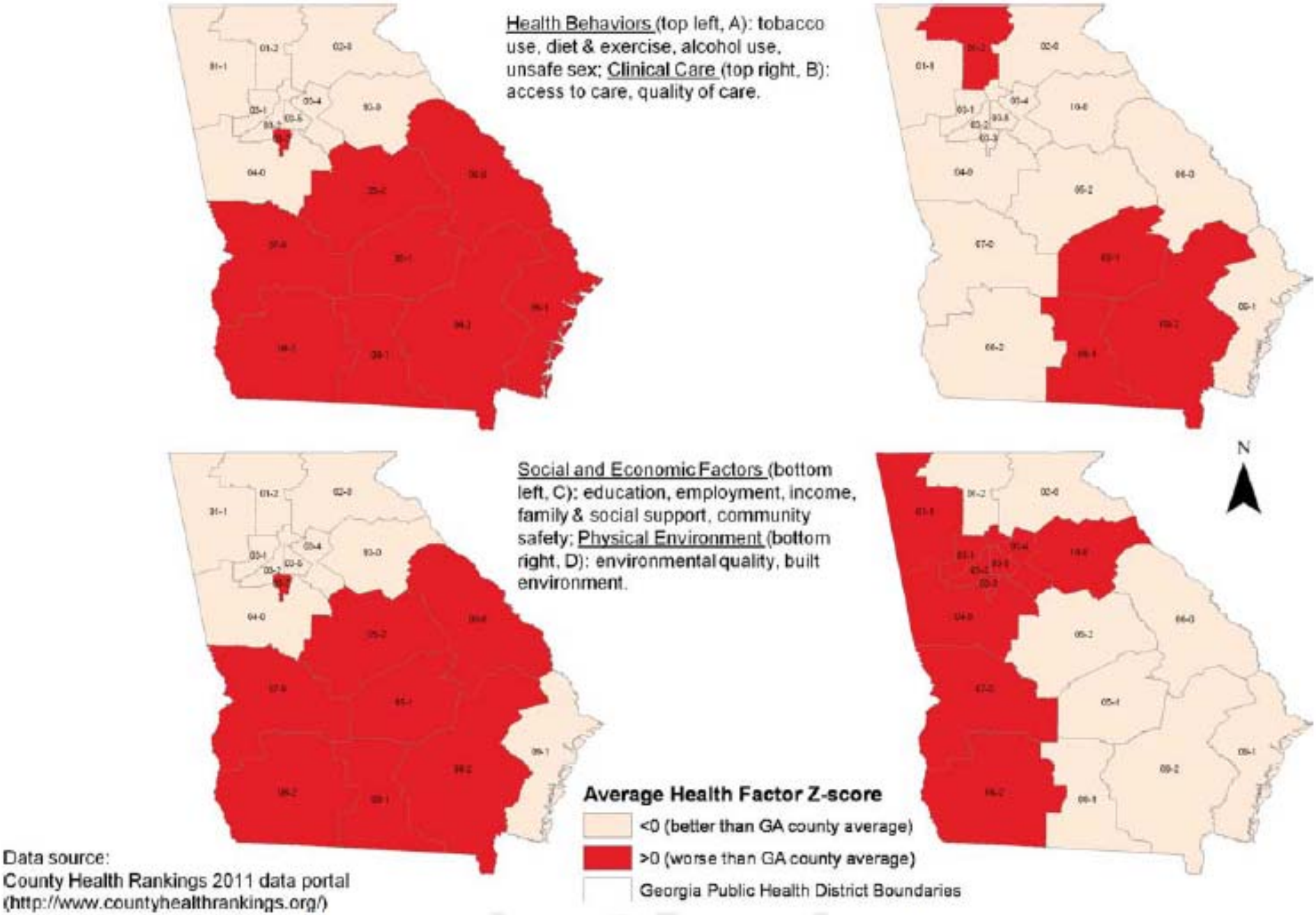


Figure 8. Average health factor scores are illustrated by Georgia public health district.

Discussion

- Blacks had more fatal cancers than whites for all cancer sites
- Higher MIRs were observed among blacks in SC compared with blacks in GA
  - Except cervical cancer
- Worst health outcomes in West and East Central health districts

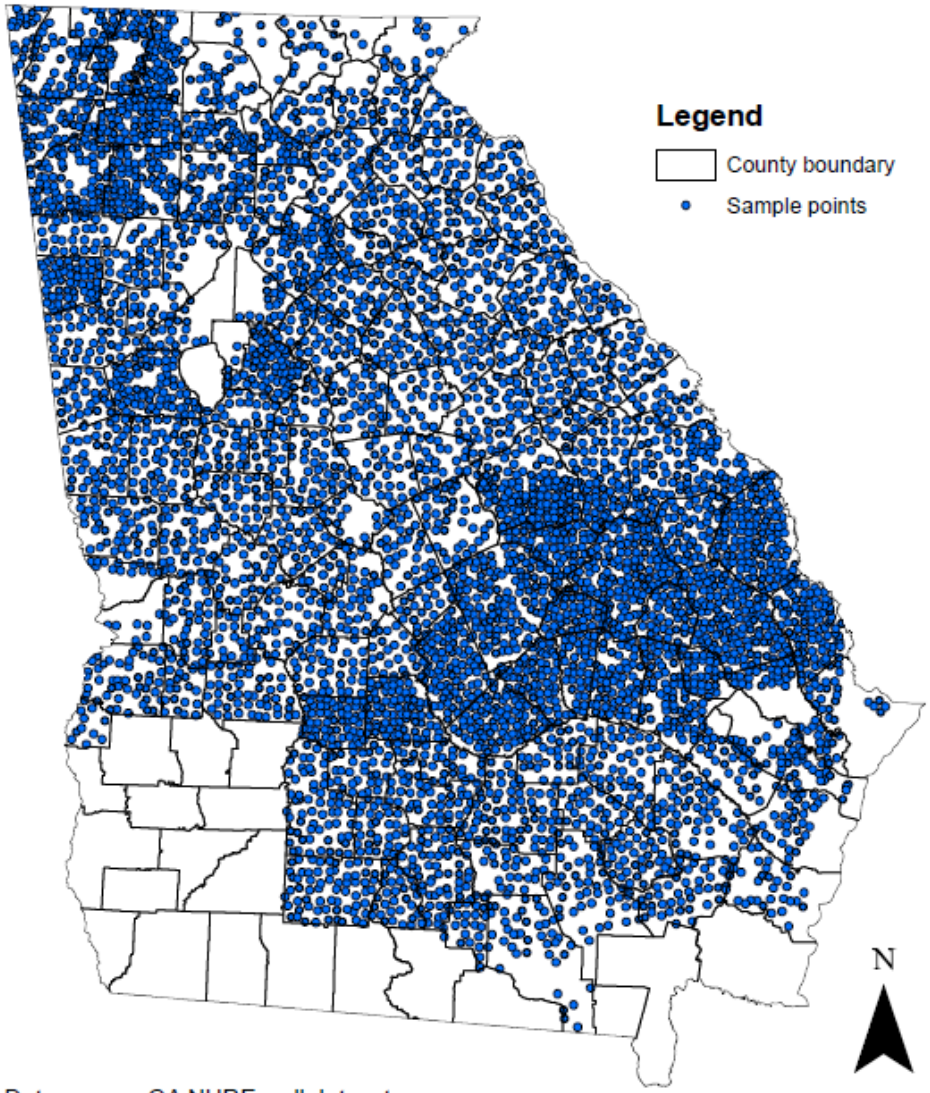
Strengths/Limitations

- County Health Rankings system
- Use of health districts
- MIR
  - Efficient
  - Avoids survival studies
- GCCR data

Summary

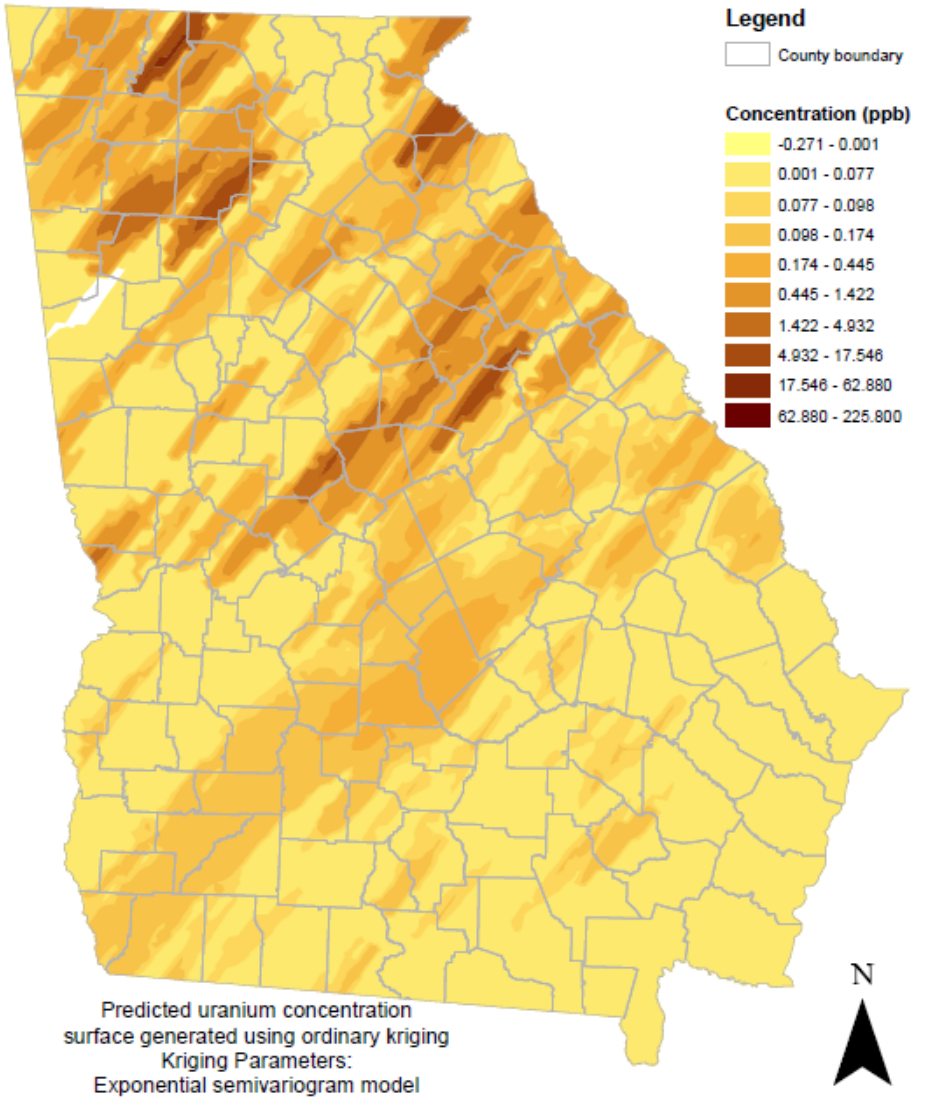
- Larger MIRs for blacks
  - Especially prostate, cervical, oral cancer in men
- More fatal cancers in West and East Central Georgia
- May be related to health behavior, clinical care, social/economic factors

Uranium Sample Points



Data source: GA NURE well dataset
N=5,633 uranium samples

Predicted Well Uranium Concentrations



Predicted uranium concentration
surface generated using ordinary kriging
Kriging Parameters:
Exponential semivariogram model
Southwest-northeast anisotropy (minor range: 0.5; direction: 40)

N=24,355 samples in bordering states
N=5,633 uranium samples

Figure 3. Predicted radon concentration surface.

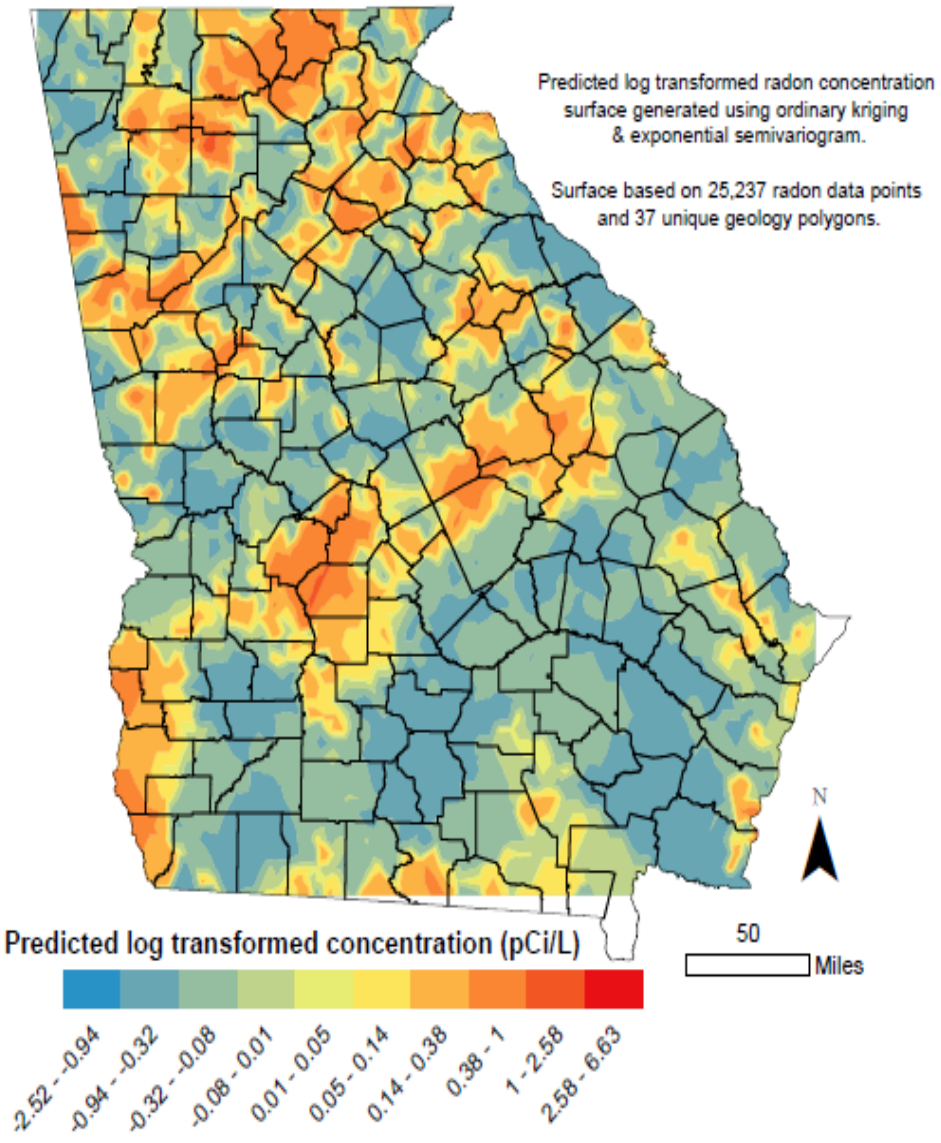


Table 1. Adjusted and stratified relationship of predicted log-transformed radon concentrations [log(pCi/L)] by county and cancer incidence (specific cancer site vs. cervical cancer control).

	Unadjusted		Adjusted*	
Cancer Site	Odds Ratio	95% CI	Odds Ratio	95% CI
Lung & bronchus (n=34,355)	1.02	0.93, 1.13	0.97	0.86, 1.08
Kidney & renal pelvis (n=6,539)	1.04	0.94, 1.16	0.87	0.76, 1.01
Female breast (n=33,540)	**1.21**	**1.10, 1.34**	1.07	0.97, 1.19
Leukemia (n=10,195)	1.11	0.97, 1.23	0.91	0.68, 1.24
Colorectal (n=23,077)	1.03	0.93, 1.14	0.98	0.89, 1.12
Urinary bladder† (n=8,228)	**1.20**	**1.07, 1.33**	0.99	0.83, 1.19

Bold values refer to a statistically significant association based on confidence intervals.
CI: confidence interval
Cervical cancer controls N=2,323.
*Adjusted models control for: race, gender (except breast), tumor stage, county-level median household income.
† Urinary bladder cancer includes *in situ* cases.

Data Layers

Cancer Data: Georgia Comprehensive Cancer Registry

Point locations of subjects with seven cancer types:

1. Bladder
2. Breast
3. Colorectal
4. Kidney
5. Leukemia
6. Lung
7. Other

Individual-level covariates:

- Age
- Gender
- Race (White vs. Non-White)

Radionuclide data

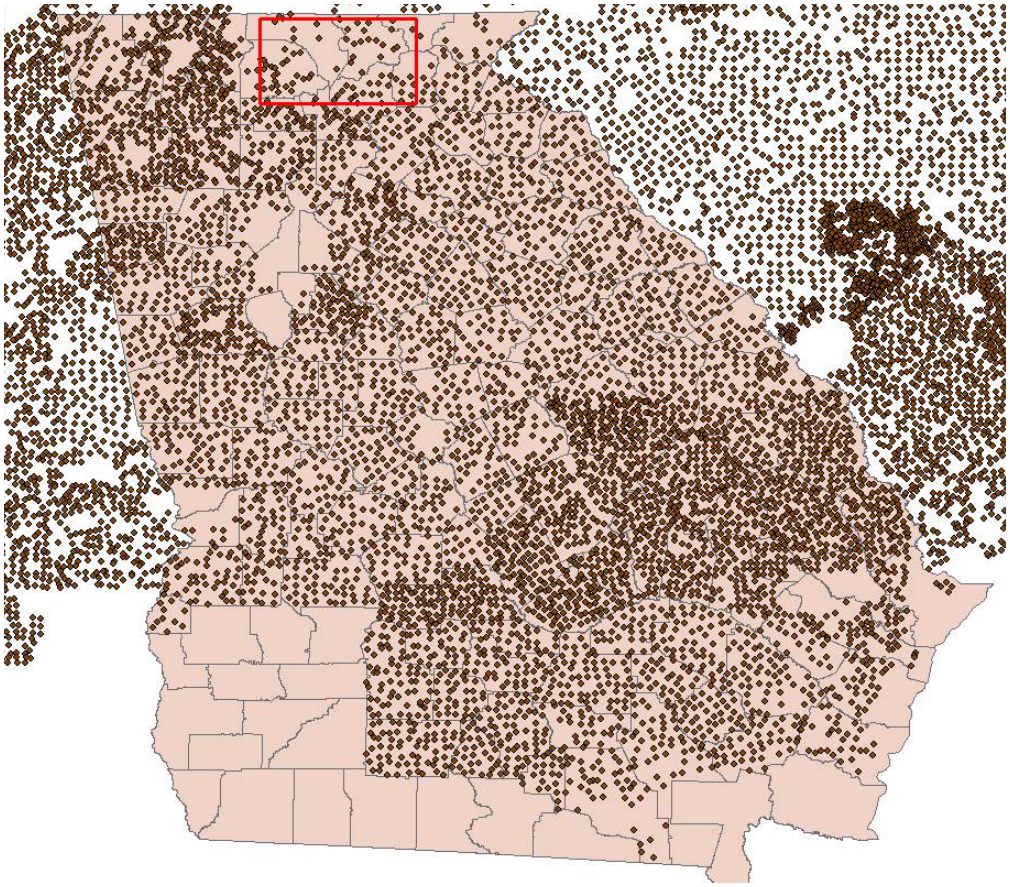
Geologic Data

Data Layers

Cancer Data: Georgia Comprehensive Cancer Registry

Radionuclide data

Groundwater Uranium. National Uranium Resource Evaluation Program



Household Radon. University of Georgia Cooperative Extension

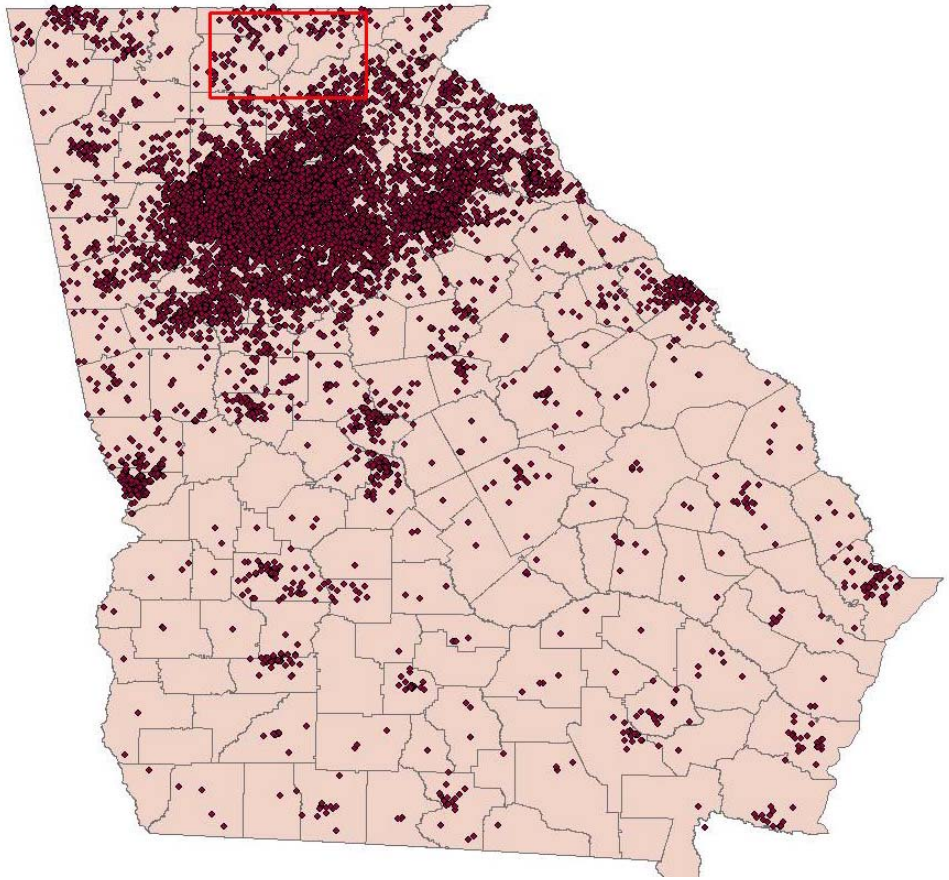
Geologic Data

Data Layers

Cancer Data: Georgia Comprehensive Cancer Registry

Radionuclide data

Groundwater Uranium. National Uranium Resource Evaluation Program
Household Radon. University of Georgia Cooperative Extension



Geologic Data

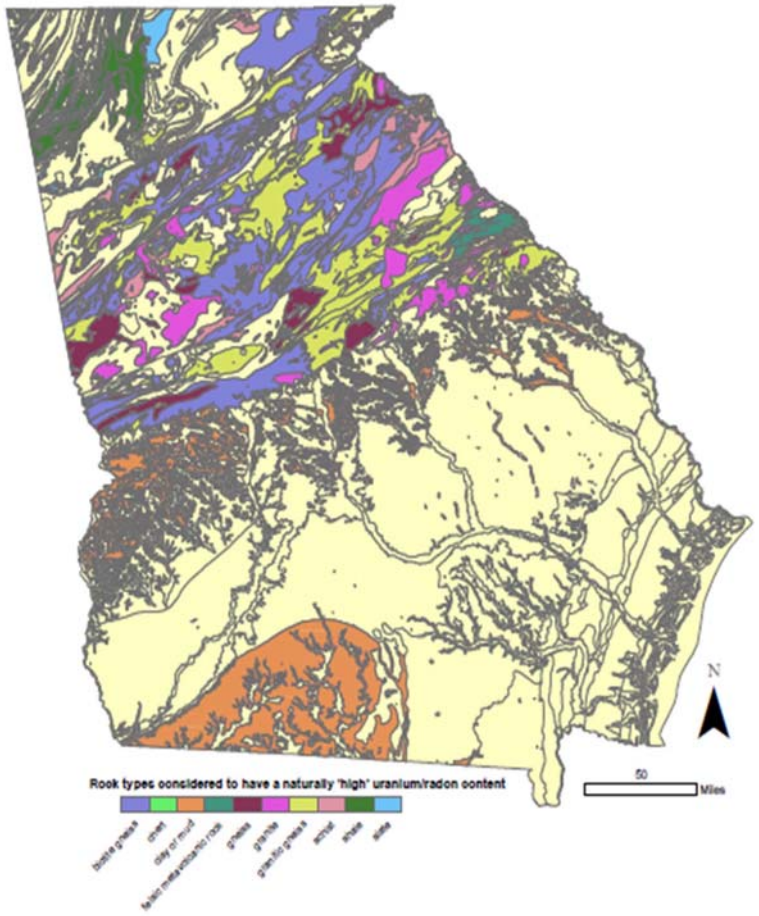
Data Layers

Cancer Data: Georgia Comprehensive Cancer Registry

Radionuclide data

Groundwater Uranium. National Uranium Resource Evaluation Program
Household Radon. University of Georgia Cooperative Extension

Geologic Data



Three-Stage Hierarchical Spatial Model

I. **Stage 1**. Geostatistical model for radionuclide levels.
II. **Stage 2**. Logistic regression model for cancer data.
III. **Stage 3**. Prior Model

Three-Stage Hierarchical Spatial Model

I. **Stage 1**. Geostatistical model for radionuclide levels.

Radionuclide level $Z(s)$ at location s is described by the spatial regression model

$$\log Z(s) = \mu_i + \varepsilon(s);\ s \in A_i,$$

where A_i is the set of locations in rock type i, μ_i is the mean radionuclide level in rock type i, $\varepsilon(s)$ is a zero-mean Gaussian random field with exponential covariance function

$$C(r;\theta) = \begin{cases} \sigma^2(1-d_0)\exp\{-\alpha r\}; & r > 0 \\ \sigma^2; & r = 1 \end{cases}$$

and r is great circle distance.

II. **Stage 2**. Logistic regression model for cancer data.

III. **Stage 3**. Prior Model

Three-Stage Hierarchical Spatial Model

I. **Stage 1**. Geostatistical model for radionuclide levels.

II. **Stage 2**. Logistic regression model for cancer data.

Conditional on the realization of the radionuclide random field:

- Controls are sampled from a point process with baseline intensity

$$\lambda_0(u)$$

- Cancer cases are sampled from a point process with intensity

$$\lambda_1(u) = \lambda_0(u)\exp\{\beta' x(u)\}$$

where the vector $x(u)$ includes log radionuclide concentration, and confounders such as age, race, etc.

If both point processes are Poisson then the cancer indicators $Y(u_i)$ for event at location u_i; $i = 1, \cdots, m$ are independently sampled from a Bernoulli distribution with probabilities $q(u_i)$ described by the logistic regression model

$$\log \frac{p(u_i)}{1 - p(u_i)} = \beta' x(u)$$

III. **Stage 3**. Prior Model

Three-Stage Hierarchical Spatial Model

I. **Stage 1**. Geostatistical model for radionuclide levels.

II. **Stage 2**. Logistic regression model for cancer data.

III. **Stage 3**. Prior Model

Geostatistical Model for Radionuclide Exposure

$$\log Z(s) = \mu_i + \varepsilon(s);\ s \in A_i$$

$$C(r;\theta) = \sigma^2(1 - d_0)\exp\{-\alpha r\};\ r > 0$$

- Mean log radionuclide levels μ_i

$$\pi(\mu_i) \propto 1$$

- Variance σ^2

$$\pi(\sigma^2) \propto 1/\sigma^2$$

- Nugget effect

$$d_0 \sim U(0,1)$$

- Range parameter

$$e^{\alpha} \sim U(0,1)$$

Handcock and Stein (1993) *Technometrics* **35**, 403-410.

Three-Stage Hierarchical Spatial Model

I. **Stage 1**. Geostatistical model for radionuclide levels.

II. **Stage 2**. Logistic regression model for cancer data.

III. **Stage 3**. Prior Model

Logistic regression model for cancer indicator

$$\log \frac{p(u_i)}{1-p(u_i)} = \beta' x(u)$$

Regression coefficients

$$\beta \sim N(0, \tau^2 I),$$

Inferential Issues

Spatial Misalignment: Sites $s_1, \cdots, s_n$ at which radionulide levels are not the same as the sites $u_1, \cdots, u_m$ at which cases and controls are observed.

Some Examples:

- Zhu, Carlin and Gelfand (2003; *Environmetrics*): Effect of ozone exposure on pediatric asthma in zip codes of Atlanta.
- Greco, Lawson, Cocchi and Temples (2005; *Environmental and Ecological Statistics*): Effect of uranium exposure on cancer incidence in zip codes of northern South Carolina.
- Fuentes, Song, Ghosh, Holland and Davis (2006; *Biometrics*): Effect of $PM_{2.5}$ on deaths due to natural and cardiovascular disease in U.S. counties.
- Smith, Zhang and Field (2007; *Statistics in Medicine*): Effect of radon exposure on leukemia in Iowa counties.

In these studies:

1. Disease counts are aggregated within regions such as zip codes or counties
2. Exposure measured at point locations.

Issue: Ecologic Bias

Bayesian Approaches

Notation:

Observed Radionuclide Data	$Z_0 = (Z(s_1), \cdots Z(s_n))'$
Unobserved Radionuclide Data at Case/Control Sites	$Z_1 = (Z(u_1), \cdots Z(u_m))'$
Cancer Indicator at Case/Control Sites	$Y = (Y(u_1), \cdots Y(u_m))'$
Radionuclide Parameters	$\theta = (\mu, \sigma^2, d_0, \alpha)$
Cancer Parameters	β

Marginal distribution of observed data:

$$p(Y, Z_0|\theta, \beta) = \int p(Y|Z_1, \beta) p(Z_1|Z_0; \theta, \alpha) p(Z_0|\theta, \alpha) dZ_1$$

Data Augmentation (Chib and Greewood 1998; Weir and Pettitt 1999, 2000; De Oliveira 2000)

Instead of drawing samples from the posterior distribution

$$p(\theta, \beta|Y, Z_0) \propto p(Y, Z_0|\theta, \beta)\pi(\theta, \beta)$$

samples are drawn from

$$p(\theta, \beta, Z_1|Y, Z_0) \propto p(Y|Z_1, \beta) p(Z_1|Z_0; \theta) p(Z_0|\theta)\pi(\theta, \beta)$$

Alternative Approaches:

- *Fully Bayesian Approach*: Sample from the posterior distribution

$$p(\theta, \beta, Z_1|Y, Z_0) \propto p(Y|Z_1, \beta)p(Z_1|Z_0; \theta)p(Z_0|\theta)\pi(\theta, \beta)$$

- *Two-Stage Bayesian Approach*: (Gryparis et al. 2009; Lee and Shadick 2010)

  Stage 1. Sample θ from the posterior of the exposure model

$$p(\theta|Z_0) \propto p(Z_0|\theta)\pi(\theta)$$

  and sample the unobserved Z_1 from the posterior predictive distribution

$$p(Z_1|Z_0; \theta)$$

  Stage 2. Sample β from the posterior of the health-effects model

$$p(\beta, Z_1|Y, Z_0, \theta) \propto p(Y|Z_1, \beta)p(Z_1|Z_0; \theta)\pi(\beta)$$

  where the posterior predictive distribution $p(Z_1|Z_0; \theta)$ is treated as the prior for Z_1 in the health-effects model.

Consider two versions of the two-stage Bayesian approach:

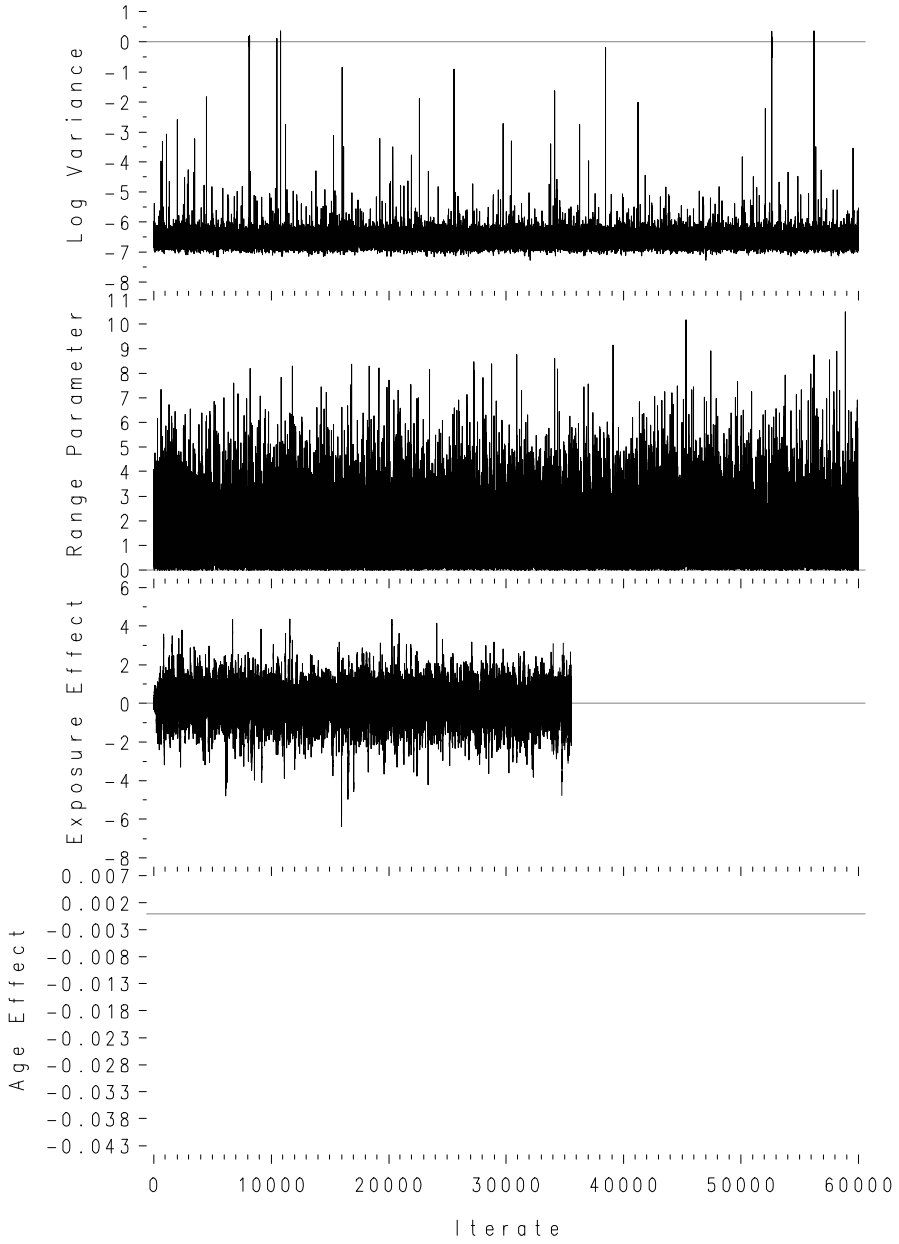
1. No feedback between health outcome and exposure estimates.
2. Yes, there is feedback between health outcome and exposure estimates.

Breast Cancer Data: Treat remaining cancer types as a control group.

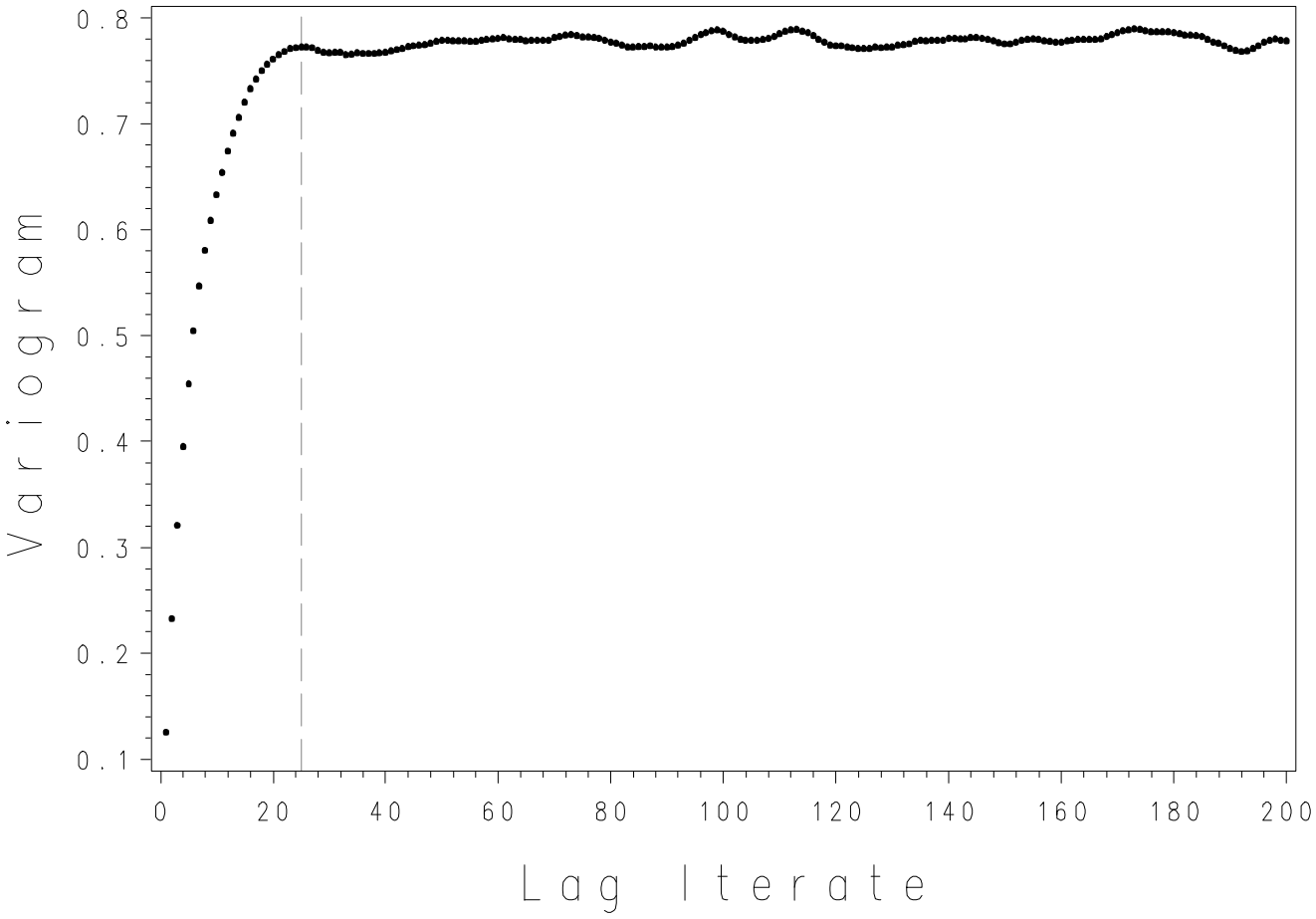
Predictors:

Exposure Model	Cancer Indicator
4 Rock Types	Uranium Exposure (well water)
	Age

Iterates of MCMC Algorithm for Breast Cancer



Variogram against lag iterate for exposure effect.



Adjusted odds ratios for the effects of uranium exposure for two-stage Bayesian inference with and without feedback between pattern of cancer cases and predicted uranium exposures.

Cancer	No Feedback	Feedback
Bladder	0.02 (0.00, 1.08)	0.01 (0.00, 1.05)
Breast	1.38 (0.14, 4.65)	1.36 (0.16, 5.06)
Colorectal	2.17 (0.40, 12.95)	2.32 (0.50, 13.69)
Kidney	0.10 (0.00, 4.08)	0.08 (0.00, 7.42)
Lung	0.84 (0.18, 3.13)	0.84 (0.14, 3.15)

Additional Challenges

- Left censoring of Radon levels below minum detection limits.

  Solution: Method of De Oliviera (2005; *Journal of Computational and Graphical Statistics*).

- Large number of observations of radon and uranium in the full Georgia data set.

  Solutions:

  - Predictive process model (Finley, Sang, Banerjee and Gelfand 2009; *Computational Statistics and Data Analysis*)
  - Fixed rank kriging (Cressie and Johannson 2008; *JRSSB*).