Report on the

Analysis of Environmental Factors and Cancer Incidence in New Jersey

Demonstration Project on

Geographic Patterns of Cancer Incidence and Environmental Factors in New Jersey
Phase 2: Epidemiologic (Linking) Studies

Program 03074, Environmental and Health Effects Tracking
National Center for Environmental Health
Centers for Disease Control and Prevention (CDC)

Submitted to CDC
January 31, 2007

Prepared by the

New Jersey Department of Health and Senior Services
Consumer and Environmental Health Services
Environmental Public Health Tracking Project
and
Cancer Epidemiology Services

In cooperation with the
New Jersey Department of Environmental Protection

Table of Contents

Overview of the Demonstration Project

The New Jersey Department of Health and Senior Services (NJDHSS) was awarded funding from the Centers for Disease Control and Prevention (CDC) to conduct three demonstration projects under the program, "Environmental and Health Effects Tracking," in cooperation with the New Jersey Department of Environmental Protection (NJDEP). The purpose of these demonstration projects is to develop and evaluate methods for linking ongoing, existing health effects and human exposure surveillance systems with existing systems for monitoring environmental hazards and exposures.

One of the three demonstration projects by NJDHSS and NJDEP is to link cancer incidence data with data on environmental hazards and exposures. Environmental factors are known or suspected to play an important role in the etiology of several cancer types. This demonstration project will allow NJDHSS and NJDEP to proactively evaluate the geographic relationships among the incidences of selected cancer types and specific environmental hazards or exposures.

The project was conducted in two phases. Phase 1, described in a separate report, involves identification of specific cancer types of interest, and descriptive analysis of incidence data for these cancers, specifically for temporal trends and spatial patterns. The second phase, described in this report, involves the linkage of the cancer incidence and environmental databases to examine specific relationships suggested by the literature.

This demonstration project was conducted by the Environmental Public Health Tracking Project (EPHT) in Consumer and Environmental Health Services, NJDHSS, in partnership with Cancer Epidemiology Services (CES), NJDHSS, and the New Jersey Department of Environmental Protection (NJDEP).

Study Team

New Jersey Department of Health and Senior Services

Consumer and Environmental Health Services

Jerald Fagliano, M.P.H., Ph.D., Principal Investigator
Michael Berry, M.P.H.
Perry Cohn, Ph.D., M.P.H.
Barbara Goun, M.P.H., Ph.D., Project Coordinator
Patricia Haltmeier, L.P.N.
Richard Opiekun, M.S., M.A., Ph.D.

Cancer Epidemiology Services
Kevin Henry, PhD.
Betsy Kohler, M.P.H., C.T.R
Lisa Roche, M.P.H., Ph.D.

New Jersey Department of Environmental Protection

Steven Anderson, M.S.
Linda Bonanno, Ph.D.
Judith Klotz, M.S., Dr.P.H., consultant

Summary

These studies examined the spatial relationship between the incidence of certain cancers and conducted linkages to selected environmental factors. Four cancer types were selected to link with environmental hazard data: leukemia; brain and other nervous system (ONS) cancers; angiosarcoma of the liver; and bladder cancer. The studies use an ecologic design, with the census tract as the geographic unit of analysis. Incident case counts of the selected cancers served as the numerator, and population counts served as the denominator in relative rate comparisons. Four selected cancers were evaluated in three different environmental studies:

- association of ambient air benzene with leukemia;
- association of ambient air vinyl chloride with angiosarcoma, or brain/ONS cancers; and
- association of trihalomethanes in community drinking water with bladder cancer.

Rate ratios by environmental exposure level were estimated using a Poisson regression model.

No association was observed between benzene in ambient air, as estimated by the 1996 USEPA National Air Toxics Assessment (NATA), and the incidence of leukemia in New Jersey, once an important source of selection bias was identified in the initial analyses. The exclusion of cases whose address could not be geocoded using a full address resulted in differential loss of cases in areas with low estimated benzene exposure, since rural areas in New Jersey have lower estimated benzene levels and higher rates of cases with addresses that cannot be precisely geocoded. This resulted in an overestimate of the relative rate comparing high-exposure to low-exposure areas. Once cases were included who could only be geocoded based on zip code centroid, the spurious associations disappeared.

An opposite selection bias situation was encountered with the analysis of vinyl chloride and brain cancer. In New Jersey, estimates of vinyl chloride levels in air are higher in certain rural areas of southern New Jersey. Consequently, there was a relatively high loss of cases in high exposure areas, resulting in an apparently spurious negative association between vinyl chloride and brain cancer incidence.

Although it is biologically plausible that there may be associations between benzene in ambient air and leukemia risk, or between vinyl chloride and brain cancer or angiosarcoma risk, this demonstration project did not find these relationships. Exposure misclassification must be considered as a possible source of bias toward the null, i.e., no association.

Associations between trihalomethanes in drinking water and bladder cancer incidence were found in both sexes. For the highest trihalomethane exposure category, the adjusted RR was 1.09 (95% confidence interval, 1.06-1.12), which is statistically significantly higher than the referent trihalomethane category. This finding is consistent with the epidemiologic literature.

Several very important lessons learned have come about as the result of these demonstration projects. Possibly the most important is an enhanced awareness regarding the potential effects of differential case loss by geographic region. We found that geocoding success rates for NJ State Cancer Registry cases varied by both time and geographical region in New Jersey. The impact of this differential geocoding case loss, which was especially high in rural areas, has potential consequences in inferences made from the analyses, especially when the environmental variable of interest also has an urban to rural gradient, as occurred with several of the linking analyses. Consequently, future EPHT linkage efforts must carefully evaluate each health outcome dataset's geographic information to determine the most appropriate geographic scale to use in order to minimize loss and assure proper linkage with environmental exposure.

Introduction

As part of the Environmental and Health Effects Tracking demonstration project on cancer and environmental factors, the NJDHSS, in cooperation with the NJDEP, sought to develop and evaluate methods for linking ongoing, existing health effects and human exposure surveillance systems with existing systems for monitoring environmental hazards and exposures.

Based on a review of evidence in the literature, and the feasibility of developing relevant environmental exposure metrics, the NJDHSS and NJDEP focused research on specific pollutant exposures, populations and specific cancer types most likely to be important for epidemiologic study. The following three study questions were selected for epidemiologic examination:

> Is there an association between benzene exposure in ambient air and the incidence of leukemias?
>
> Is there an association between vinyl chloride in ambient air and the incidence of a) brain and other nervous system cancers or b) angiosarcoma of the liver?
>
> Is there an association between disinfection by-products in drinking water and the incidence of bladder cancer?

For each of these study questions, there is either sufficient or suggestive evidence of an association between exposure to the environmental factor and subsequent increased risk of the development of the cancer type. Exposure to benzene has been associated with leukemia incidence in the occupational setting (Brett, Rodricks and Chinchilli, 1989). Exposure to vinyl chloride has been associated with liver angiosarcoma and possibly brain and other nervous system cancers also in the occupational setting (McLaughlin and Lipworth, 1999). Trihalomethanes in drinking water have been associated with the incidence of bladder cancer in a series of ecological and case-control epidemiologic studies (Freedman et al., 1997; Cantor et al., 1987; McGeehin et al., 1993; Villanueva et al., 2003; Villanueva et al., 2004).

The methodologies developed in these studies, and the experiences of linking and conducting epidemiologic analyses, are intended to form a basis for future activities for additional cancers and the other health outcomes of interest under the overall Environmental Public Health Tracking program of CDC.

Methods

Overview of Methods

These studies examine the spatial relationship between the incidence of certain cancers and potential for exposure to selected environmental factors. The studies use an

ecologic design, with the census tract as the geographic unit of analysis. Incident case counts will serve as the numerator and population counts will serve as the denominator in relative rate comparisons. Rate ratios by environmental exposure level are estimated using a Poisson regression model.

Population

The study population consists of all residents of the State of New Jersey for the period 1979 through 2002. Populations were aggregated at the census tract level. Year 2000 census tract boundary definitions were used for the study. Population estimates by age group and sex are available for the years 1980, 1990 and 2000 (GeoLytics, 2005). For the year 2000, there were 1,944 census tracts in New Jersey.

Cancer Data

The source of cancer incidence data was the New Jersey State Cancer Registry (NJSCR) within the NJDHSS. Cases identified at autopsy or via death certificate were excluded because of lack of address information and uncertainty about the address of residence at diagnosis. The residential address at time of cancer diagnosis for all cases has been geocoded by the NJSCR to the year 2000 census tracts. Approximately 92% of the non-death certificate/autopsy cases in the NJSCR database are geocoded to the census tract based on a street level address match, and much of the remainder is geocoded to the centroid of the zip code.

Four cancer types were selected to link with environmental hazard data: leukemia, brain and other nervous system (ONS) cancers, angiosarcoma of the liver, and bladder cancer.

Exposure Assessment

A summary of the development of environmental exposure metrics used for these analyses are provided below. A detailed description of the methods used to derive the metrics are found in a separate report (NJDEP, 2007).

Benzene in ambient air: Two approaches to ambient benzene exposure assessment were used. First, National Air Toxics Assessment (NATA) estimates by census tract were used as the benzene exposure metric. NATA is a metric which integrated the contribution of large point emission sources, smaller (area) sources, mobile sources, and background levels, all of which contribute to the overall metric. Modeled NATA estimates for average annual benzene concentration during the year 1996 and are available from the USEPA at the census tract level for the entire state of New Jersey. Figure 1 presents summary statistics and a frequency distribution of the benzene estimates showing the cut points used in the analysis. All census tracts in the state have a concentration exceeding the one-in-a-million cancer risk estimate of 0.13 micrograms per cubic meter (ug/m^3). The maximum modeled concentration was 4.57 ug/m^3. Exposure variable cut points used in the analysis were chosen as 1.4 ug/m^3 and 2.5 ug/m^3.

The second approach utilized facility emission databases to construct a benzene exposure metric incorporating both point and area emission sources, using methods that may have finer geographic resolution than in NATA. The NJDEP developed a metric of benzene exposure potential based solely on facility emission and location databases for each census tract across the state. Facilities included those that are permitted for or report releases of benzene, such as petroleum refineries, chemical manufacturers, and gasoline stations. First, NJDEP developed an emission inventory of pollutant sources. Second, NJDEP conducted air dispersion modeling to estimate ambient concentrations for common source types. The third step was to overlay and integrate air dispersion results from all sources at the grid and then the census tract level.

Figure 2 presents a summary of the NJDEP benzene estimates and the cut points used in the analysis. The highest concentrations were near the five petroleum refineries in New Jersey. The maximum estimated concentration from facility sources was 2.26 g/m^3. (Note that these estimates do not include contributions from mobile sources or background that are included the NATA estimates.) Concentrations of 0.13 ug/m^3 and 0.013 ug/m^3 were selected as cut points for this analysis.

Vinyl chloride in ambient air: As with benzene, the same two approaches to exposure assessment were used. NATA estimates by census tract were used as the vinyl chloride exposure metric. For vinyl chloride, large point sources are the major contributor to local variation. Figure 3 presents a summary of the NATA vinyl chloride estimates and the cut points used in the analysis. Higher concentrations are found in the southwest part of the state in Salem, Cumberland, and Gloucester Counties near a well known source of vinyl chloride air release. The maximum vinyl chloride concentration was estimated to be slightly below the one-in-a-million excess cancer risk concentration of 0.11 ug/m^3. Two separate two-level (i.e., high vs. low) exposure variables were developed for the 1996 NATA vinyl chloride estimates. The first exposure variable cut point was 0.01 $\mu g/m^3$, which is an approximate point of inflection in the frequency distribution, and the second exposure variable cut point was 0.005 ug/m^3, making two categories about equal in the number of census tracts.

NJDEP used the same procedure to model vinyl chloride as was done for benzene. Figure 4 presents a summary of the NJDEP vinyl chloride estimates. The maximum predicted concentration was 0.404 ug/m^3. Only four census tracts have predicted concentrations exceeding the one-in-a-million cancer risk level. The same cut points discussed for the NATA vinyl chloride exposure variables were used for the NJDEP modeled data.

Trihalomethanes (THMs) in drinking water: The NJDHSS and the NJDEP developed a metric of potential exposure to total trihalomethanes, the most common disinfection by-product in drinking water exposure, for each census tract in the state. Using GIS, NJDHSS assigned a community drinking water system code (and its accompanying water quality information) to each census tract, based on the overlay between the GIS shapefile of drinking water system boundaries and the GIS shapefile of

census tracts. The Spatial Analyst function was used to estimate weighted average exposure concentrations in census tracts served by more than one water system.

The THM data used in the analysis were estimated averages in the period 1978 to 1985, and are summarized in Figure 5. The maximum average THM level was 124.1 ug/l, 55% higher than the maximum allowable amount of 80 ug/l. A four level analysis variable was created using cut points at the inflection points of 0 ug/l, 45 ug/l, and 65 ug/l. For this analysis, populations not served by public water systems were presumed to have private wells and were assigned a value of 0 for THM concentration.

Linking Studies

The four selected cancers were evaluated in three different environmental studies, and include:

1. association of ambient air benzene with leukemia,
2. association of ambient air vinyl chloride with angiosarcoma or brain/ONS cancer, and
3. association of trihalomethanes in community drinking water with bladder cancer.

Data Analysis

We constructed an analytical dataset in which the census tract is the unit of analysis. The dataset contained case counts and population estimates for each census tract, by sex and age group. Exposure metrics were also assigned to each census tract.

Cancer incidence rate ratios (RRs) were computed for each cancer type for levels of exposure metrics. Epidemiologic analyses were conducted using Stata statistical software (StataCorp, 2003). Rate ratio estimates were computed using the Poisson regression model (Clayton and Hills, 1993). Confidence intervals (95%) were generated for the RR estimates. RRs were adjusted for age group, sex, the percent of the population below the poverty level, and the percent of the population which is white. Select sex-specific analyses were also conducted.

Evaluation of Cases Geocoded using Partial Address Information

Cases that could not be geocoded to a census tract using the entire address at time of diagnosis were dropped from the initial analyses. However, geographic analysis of non-geocoded cases revealed a sharp geographic pattern of loss: there was a greater loss in more rural/suburban areas. Therefore, these cases were subsequently geocoded to the zip centroid census tract and added to the other data for reanalysis in order to evaluate the impact of loss.

Results

Cancer Incidence Summary

Table 1 presents the case eligibility for the selected cancers. A total of 21,861 cases (12,240 males and 9,621 females) met the case definition for leukemia during the 24-year period. A total of 36 cases (20 males and 16 females) met the case definition for angiosarcoma of the liver. A total of 12,394 cases (6,755 males and 5,639 females) met the case definition for brain/ONS cancers. A total of 47,590 cases (34,714 males and 12,876 females) met the case definition for bladder cancer.

Table 2 presents the geocode success rate for the eligible cases. For leukemia, 20,130 cases (11,301 males and 8,829 females), about 92% of the cases, were geocoded to a census tract using the full address. An additional 1,579 leukemia cases (853 males and 726 females), just over 7% of the cases, were geocoded to a census tract using the zip centroid. The remaining 152 leukemia cases (85 males and 67 females), less than 1% of the cases, could not be geocoded to a census tract due to insufficient address information.

All of the angiosarcoma cases were geocoded using the full address.

For brain/ONS cancers, 11,428 cases (6,199 males and 5,229 females), about 92% of the cases, were geocoded to a census tract using the full address. An additional 918 brain/ONS cancer cases (530 males and 388 females), between 7% and 8% of the cases, were geocoded to a census tract using the zip centroid. The remaining 48 brain/ONS cancer cases (26 males and 22 females), less than 1% of the cases, could not be geocoded to a census tract due to insufficient address information.

For bladder cancer, 43,653 cases (31,877 males and 11,776 females), about 92% of the cases, were geocoded to a census tract using the full address. An additional 3,617 bladder cancer cases (2,603 males and 1,014 females), between 7% and 8% of the cases, were geocoded to a census tract using the zip centroid. The remaining 320 bladder cancer cases (234 males and 86 females), less than 1% of the cases, could not be geocoded to a census tract due to insufficient address information.

Ambient Air Benzene with Leukemia

1996 NATA benzene estimates: The breakdown of census tracts by the NATA benzene exposure variable was as follows: 852 (44%) in the referent category (less than 1.40 $\mu g/m^3$); 881 (45%) in middle exposure category (1.40 $\mu g/m^3$ to less than 2.50 $\mu g/m^3$); and 217 (9%) in high exposure category (2.50 $\mu g/m^3$ and above).

The initial analysis evaluated only cases geocoded with an entire address and is presented in Table 3. The adjusted RR for the middle NATA benzene exposure category was 1.09 (95% confidence interval, 1.06-1.12), statistically significantly higher than benzene referent exposure category. For the highest benzene exposure category, the adjusted RR was 1.04 (95% confidence interval, 0.98-1.09). The sex-specific analyses,

Tables 4a and 4b, found statistically significantly elevated adjusted RRs for females in the middle category (RR=1.08; 95% confidence interval, 1.03-1.13) and highest category (RR=1.11; 95% confidence interval, 1.03-1.21) exposure categories. The results for males more closely paralleled the analysis with the combined sexes, the middle benzene exposure category was statistically significantly elevated (RR=1.10; 95% confidence interval, 1.05-1.14), while the highest exposure category was not elevated (RR=0.97; 95% confidence interval, 0.90-1.05).

As noted above, there was considerable difference in the loss of cases eliminated from the analysis because the address could not be geocoded using full address information, and there was concern that this pattern of loss might be related to the pattern of exposure to benzene in ambient air. Because of this geographic skewing of cases, leukemia was then reanalyzed including cases that were geocoded to the zip centroid. Results from the new analyses showed that all the adjusted RRs were attenuated toward 1.0, with no exposure categories statistically significantly elevated. For both sexes combined (Table 5), the adjusted RR for the middle exposure category was 1.03 (95% confidence interval, 0.99-1.06) and for the highest exposure category was 0.97 (95% confidence interval, 0.92-1.03). For females (Table 6a), the RR for the middle exposure category was 1.03 (95% confidence interval, 0.99-1.08) and for the highest exposure category was 1.05 (95% confidence interval, 0.97-1.13). For males (Table 6b), the RR for the middle exposure category was 1.02 (95% confidence interval, 0.99-1.06) and the highest exposure category was 0.91 (95% confidence interval, 0.85-0.98), statistically significantly lower than the referent category.

NJDEP benzene modeled estimates: The breakdown of census tracts by the NJDEP benzene exposure variable was as follows: 1,112 (57%) in the referent category (less than 0.013 $\mu g/m^3$); 817 (42%) in middle exposure category (0.013 $\mu g/m^3$ to less than 0.13 $\mu g/m^3$); and 21 (1%) in high exposure category (0.13 $\mu g/m^3$ and above).

The initial analysis evaluated only cases geocoded with an entire address and is presented in Table 7. The adjusted RR for the middle NJDEP benzene exposure category was 1.07 (95% confidence interval, 1.04-1.11), statistically significantly higher than benzene referent exposure category. For the highest benzene exposure category, the adjusted RR was 1.02 (95% confidence interval, 0.88-1.18).

The reanalysis including zip centroid geocoded cases is presented in Table 8. The adjusted RR for the middle NJDEP benzene exposure category was 1.01 (95% confidence interval, 0.98-1.04). For the highest benzene exposure category, the adjusted RR was 0.97 (95% confidence interval, 0.84-1.11).

Ambient Vinyl Chloride with Angiosarcoma and Brain/ONS Cancer

1996 NATA vinyl chloride estimates: The breakdown of census tracts by the first NATA vinyl chloride cut point (0.01 $\mu g/m^{3)}$ was 1,816 (93%) in the lower exposure category and 134 (7%) in the higher exposure category. For the second vinyl chloride cut point (0.005 $\mu g/m^{3)}$, the breakdown of census tracts was 1,450 (73%) in the lower

exposure category and 520 (27%) in the higher exposure category.

NJDEP vinyl chloride modeled estimates: The breakdown of census tracts by the first NJDEP vinyl chloride cut point (0.01 μg/m$^{3)}$ was 1,888 (97%) in the lower exposure category and 62 (3%) in the higher exposure category. For the second vinyl chloride cut point (0.005 μg/m$^{3)}$, the breakdown of census tracts was 1,693 (87%) in the lower exposure category and 257 (13%) in higher exposure category.

Angiosarcoma analysis: A summary of the results of analyses using the NATA exposure metric is presented in Tables 9a and 9b. The adjusted RR for the first higher exposure NATA vinyl chloride category (0.01 μg/m^3 cut point) was 0.45 (95% confidence interval, 0.06-3.28). For the second higher exposure NATA vinyl chloride category (0.005 μg/m^3 cut point), the adjusted RR was 2.35 (95% confidence interval, 1.21-4.58), statistically significantly higher than the referent category.

A summary of the analysis results using the NJDEP modeled metric is presented in Tables 10a and 10b. The adjusted RR for the higher exposure NJDEP vinyl chloride category (0.01 μg/m^3 cut point) was 1.29 (95% confidence interval, 0.18-9.50). For the second higher exposure NJDEP vinyl chloride category (0.005 μg/m^3 cut point), the adjusted RR was 1.06 (95% confidence interval, 0.37 -2.99).

Brain/ONS cancer analysis: A summary of the analysis results using cases only geocoded using the full address are presented in Tables 11a and 11b. For the first NATA vinyl chloride variable (0.01 μg/m^3 cut point), the adjusted RR for the higher exposure category was 0.92 (95% confidence interval, 0.84-0.99), statistically significantly lower than the referent category. For the second higher exposure NATA vinyl chloride category (0.005 μg/m^3 cut point), the adjusted RR was 0.95 (95% confidence interval, 0.90-0.99), statistically significantly lower than the referent category.

A summary of the reanalysis results including zip centroid geocoded cases are presented in Tables 12a and 12b. For the first NATA vinyl chloride variable (0.01 μg/m^3 cut point), the adjusted RR for the higher exposure category was 0.98 (95% confidence interval, 0.91-1.06) relative to the referent category. For the second higher exposure NATA vinyl chloride category (0.005 μg/m^3 cut point), the adjusted RR was 0.97 (95% confidence interval, 0.93-1.01).

Since the initial and reanalysis results of the NJDEP vinyl chloride exposure with brain/ONS cancer were identical, Tables 13a and 13b present the results of the reanalysis using cases geocoded using either the full address or zip code centroid. The adjusted RR for the higher exposure NJDEP vinyl chloride category (0.01 μg/m^3 cut point) was 1.00 (95% confidence interval, 0.89-1.12). For the second higher exposure NJDEP vinyl chloride category (0.005 μg/m^3 cut point), the adjusted RR was 0.99 (95% confidence interval, 0.94 -1.05).

Trihalomethanes in Drinking Water with Bladder Cancer

The breakdown of census tracts by THM exposure was as follows: 905 (46%) in the no exposure referent category (0 µg/l); 428 (22%) in the second exposure category (greater than 0 µg/l to 45 µg/l); 452 (23%) in the third exposure category (greater than 45 µg/l to 65 µg/l); and 165 (8.5%) in the highest exposure category (greater than 65 µg/l).

A summary of the analysis results is presented in Table 14. The initial analysis evaluated only cases geocoded with an entire address. The adjusted RR for the second THM exposure category was 0.99 (95% confidence interval, 0.96-1.01). For the third THM exposure category, the adjusted RR was 1.04 (95% confidence interval, 1.02-1.07), statistically significantly higher than the referent THM category. For the highest THM exposure category, the adjusted RR was 1.15 (95% confidence interval, 1.11-1.18), statistically significantly higher than the referent THM category.

Bladder cancer was reanalyzed using both the initial analysis cases and the zip centroid geocoded cases. A summary of the reanalysis results is presented in Table 15. The adjusted RR for the second THM exposure category was 0.96 (95% confidence interval, 0.94-0.98), statistically significantly lower than the referent THM category. For the third THM exposure category, the adjusted RR was 0.98 (95% confidence interval, 0.96-1.01). For the highest THM exposure category, the adjusted RR was 1.09 (95% confidence interval, 1.06-1.12), statistically significantly higher than the referent THM category. Sex-specific analysis showed that the increased risk in the highest exposed category was consistent for both males and females.

Discussion

Interpretation of Findings

No association was observed between benzene in ambient air, as estimated by the 1996 USEPA NATA, and the incidence of leukemia in New Jersey, once an important source of selection bias was identified in the initial analyses. The exclusion of cases whose address could not be geocoded using a full address resulted in differential loss of cases in areas with low estimated benzene exposure, since rural areas in New Jersey have lower estimated benzene levels and higher rates of cases with addresses that cannot be exactly geocoded. This resulted in an overestimate of the relative rate comparing high-exposure to low-exposure areas. Once cases were included who could only be geocoded based on zip code centroid, the spurious associations were attenuated.

An opposite selection bias situation was encountered with the analysis of vinyl chloride and brain cancer. In New Jersey, estimates of vinyl chloride levels in air are higher in certain rural areas of southern New Jersey. Consequently, there was a relatively high loss of cases in high exposure areas, resulting in an apparently spurious negative association between vinyl chloride and brain cancer incidence.

Although it is biologically plausible that there may be associations between benzene in ambient air and leukemia risk, or between vinyl chloride and brain cancer or angiosarcoma risk, this demonstration project did not find relationships. Exposure misclassification must be considered as a possible explanation, since errors of this type will result in bias toward the null, i.e., no association.

The number of angiosarcoma cases is very low (36 cases in a 24-year period) reducing statistical power of any analysis, and small changes in exposure classification may result in large changes in relative risk between exposure models. The change in relative risk from 2.4 for the NATA vinyl chloride variable (0.005 $\mu g/m^3$ cut point) to 1.1 for the NJDEP vinyl chloride variable reflects a shift of three cases from the higher to lower exposure category.

Another issue to consider is the degree to which the exposures vary across the state, and how those exposures compare to the estimated cancer risk. While there is a sevenfold spread in the NATA benzene exposure statewide, all areas of the state have an estimated level above the health benchmark. For the NATA vinyl chloride estimate, however, all of the state is below the health benchmark, and most of New Jersey is tenfold lower.

Associations between trihalomethanes in drinking water and bladder cancer incidence were found in both sexes. This finding is consistent with the epidemiologic literature.

Lessons Learned from the Demonstration Project

This demonstration project resulted in a successful collaboration among staff of the EPHT project in CEHS, Cancer Epidemiology Services, and NJDEP. Through the development of a protocol, agency representatives were able to define the questions and to design the exposure assessments and analytical approaches used in the study.

Exposure misclassification in linkage studies, as with any epidemiologic study, is an important consideration. There are many sources of error in using the NATA or drinking water data for human exposure assessment. For this reason, the results of ecologic linkage studies must be interpreted cautiously.

A significant issue in linkage studies for EPHT is the quality and fineness of scale for geocoding of health outcome data. The general desire when conducting geocoding is to code to the smallest geographic scale possible, such as a latitude/longitude location. However, the inability to geocode some percentage of health outcome data to below the municipal level means that these data are either lost from the analysis or must be estimated based on partial address information. In this evaluation, about 92% of the cancer data could be geocoded to the census tract with relative confidence. Much of the rest of the data could be assigned a census tract using partial address information (i.e., less confidently geocoded).

However, importantly, loss of case data is differential around the state. A strong geographic bias in the loss of data for accurate geocoding is evident in the cancer data from the New Jersey State Cancer Registry (NJSCR) due to a lack of adequate address information. The loss of cases due to this problem is much much greater for more rural counties in the state (NJDHSS 2007). The loss is also more severe in the earlier years of the NJSCR, decreasing towards the present time. The impact of case loss has potential consequences in inferences made from the analyses. This is especially true when the environmental variable of interest also has an urban to rural gradient, as seen with several of the linking analyses. Spurious associations were originally seen for benzene and leukemia (positive) and vinyl chloride and brain cancer (negative) due to this form of selection bias. Consequently, future data linking efforts must evaluate the health data's geographic information to determine the most appropriate geographic scale in order to minimize loss and maximize the exposure assessment.

References

Brett, SM, Rodricks, JV, Chinchilli, VM (1989). Review and update of leukemia risk potentially associated with occupational exposure to benzene. Environmental Health Perspectives 82: 267-281.

Cantor, KP., Hoover, R., Hartge, P., Mason, TJ., Silverman, DT., Altman, R., Austin, DF, Child, MA., Key, CR., Marrett, LD., Myers, MH., Narayana, AS., Levin, LI., Sullivan, JW., Swanson, GM., Thomas, DB., West, DW. (1987). Bladder cancer, drinking water source, and tap water consumption: a case-control study. Journal of the National Cancer Institute 79(6): 1269-1279.

Clayton D and Hills M. Statistical Models in Epidemiology. Oxford University Press; 1st edition. 1993.

Freedman, DM., Cantor, KP., Lee, NL., Chen, LS., Lei, HH., Ruhl, CE., Wang, SS. (1997). Bladder cancer and drinking water: a population-based case-control study in Washington County, Maryland (United States). Cancer Causes and Control 8: 738-744.

GeoLytics, Inc. CensusCD Neighborhood Change Database (NCDB) 1970–2000 US Census Tract Data [computer software]. East Brunswick, NJ, 2005.

McGeehin, MA., Reif, JS., Becher, JC., Mangione, EJ. (1993). Case-control study of bladder cancer and water disinfection methods in Colorado. American Journal of Epidemiology 138(7): 492-501.

McLaughlin, JK, Lipworth, L. (1999). A critical review of the epidemiologic literature on health effects of occupational exposure to vinyl chloride. Journal of Epidemiology and Biostatistics 4(4): 253-275.

NJDEP. Report of the Development and Evaluation of Environmental Exposure Metric Databases. Trenton, NJ, 2007.

NJDHSS. Report on the Spatial and Temporal Variation in the Incidence of Selected Cancers. Demonstration Project on Geographic Patterns of Cancer Incidence and Environmental Factors in New Jersey Phase 1: Surveillance (Tracking) Activities. Trenton, NJ, 2007.

StataCorp. Stata Statistical Software: Release 8. College Station, TX, 2003.

Villanueva, CM., Fernandez, F., Malatas, N., Grimalt, JO, Kogevinas, M. (2003). Meta-analysis of studies on individual consumption of chlorinated drinking water and bladder cancer. Journal of Epidemiology and Community Health 57(3): 166-173.

Villanueva, CM., Cantor, KP., Cordier, S., Jaakkola, JJ., King, WD., Lynch, CF., Porru, S., Kogevinas, M. (2004). Disinfection by-products and bladder cancer: a pooled

analysis. Epidemiology 15(3): 357-367.

Figures and Tables

Figure 1: Benzene Air Concentrations, NATA 1996

Descriptive Measure	Benzene (ug/m3)
Mean	1.6359
Median	1.5000
Standard Deviation	0.6618
Standard Error	0.0150
Minimum	0.6490
Maximum	4.5700

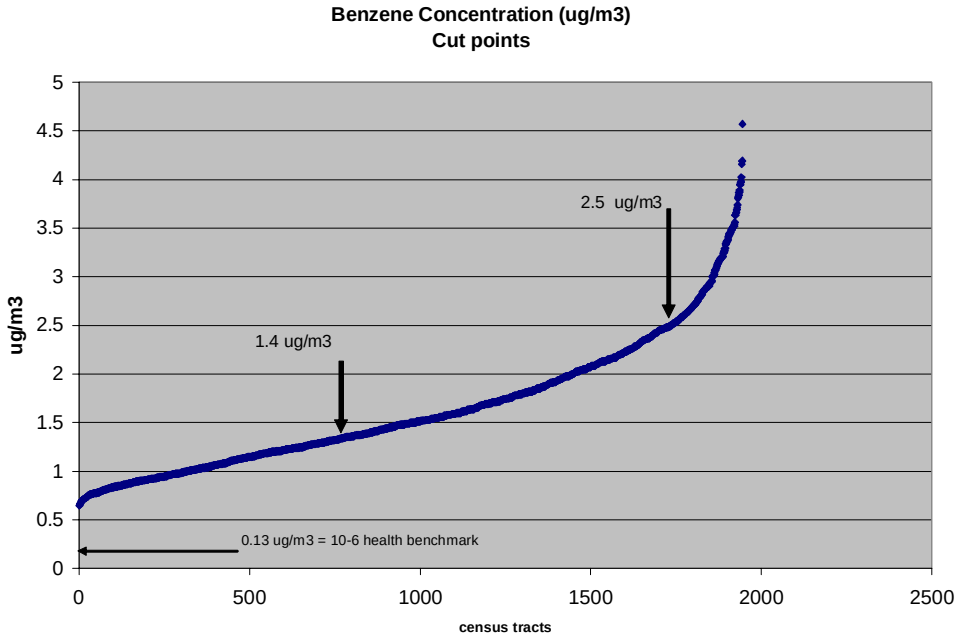

Figure 2: Benzene Air Concentrations, NJDEP 1993 – 2004

Descriptive Measure	Benzene
Mean	0.01998
Median	0.01026
Standard Error	0.00131
Standard Deviation	0.05795
Sample Variance	0.00336
Minimum	0.00000
Maximum	2.26491

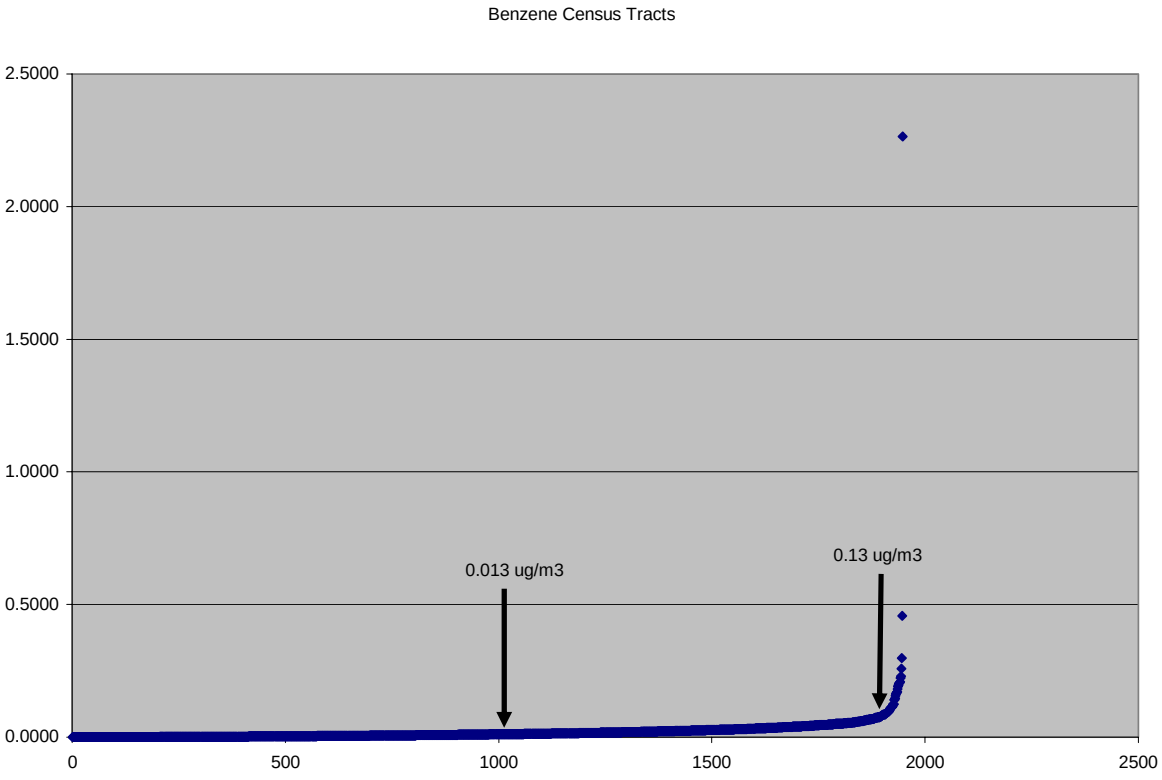

Figure 3: Vinyl Chloride Air Concentrations, NATA 1996

Descriptive Measure	VCM (ug/m3)
Mean	0.0049
Median	0.0033
Standard Error	0.0002
Standard Deviation	0.0074
Minimum	0.0001
Maximum	0.1050

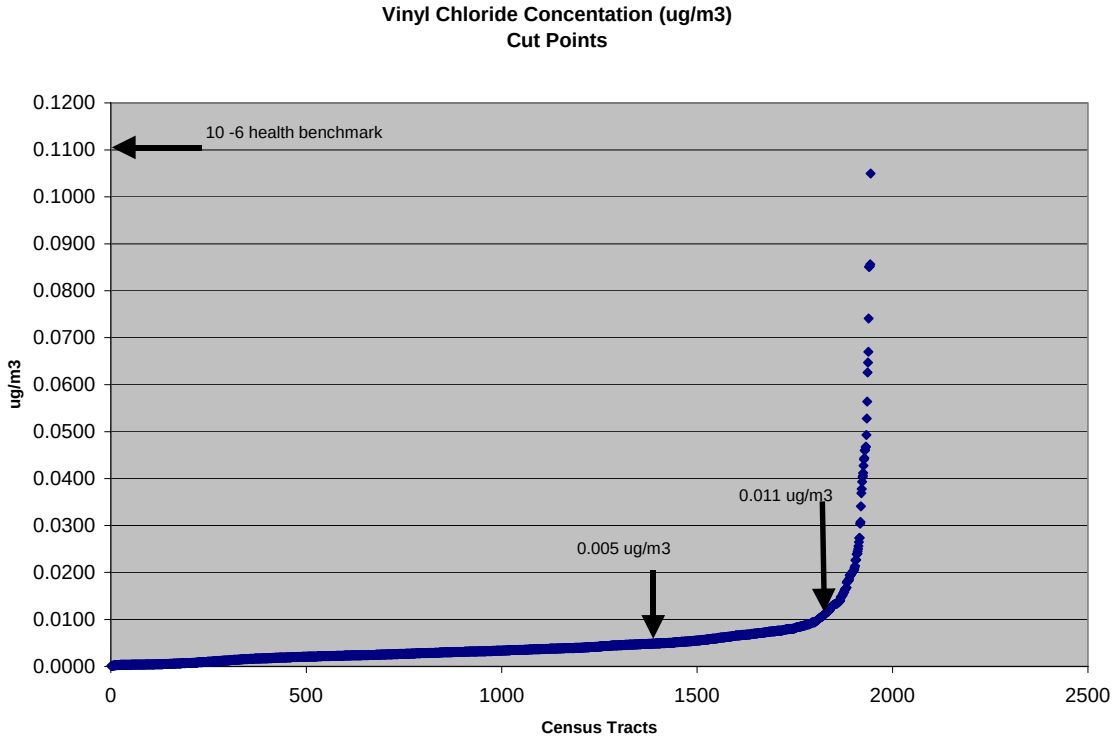

Figure 4: Vinyl Chloride Air Concentrations, NJDEP 1993 – 2004

Descriptive Measure	VCM (ug/m3)
Mean	0.00241
Median	0.00002
Standard Deviation	0.01364
Standard Error	0.00031
Minimum	0.00000
Maximum	0.40396

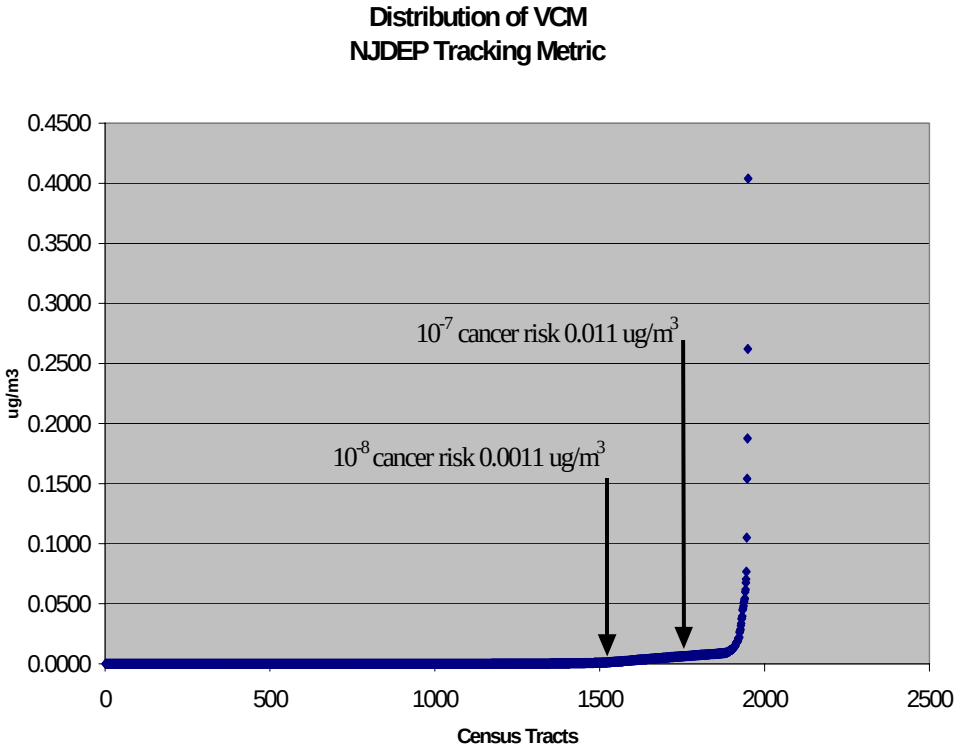

Figure 5: Total Trihalomethanes (THM) in Drinking Water, NJDEP 1978 – 1985

Descriptive Measure	THM ug/l
Mean	25.28
Median	7.70
Standard Error	0.64
Standard Deviation	28.28
Minimum	0.00
Maximum	124.10

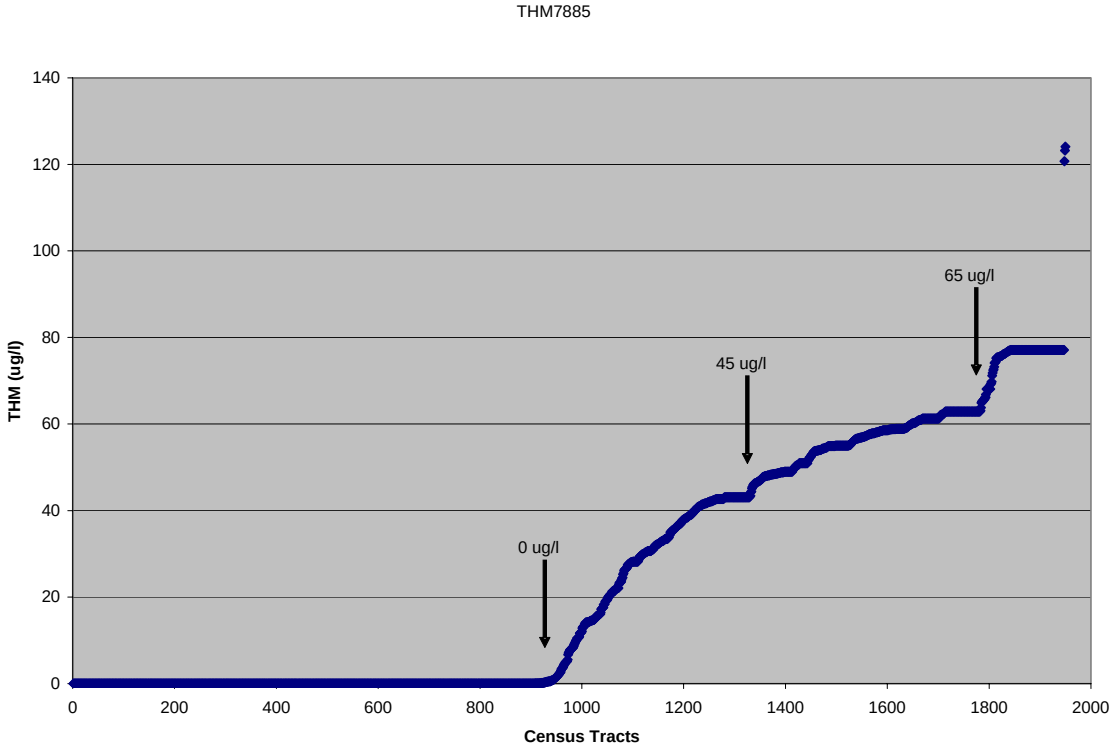

Table 1. Eligibility of Cases for Selected Cancers, 1979-2002

Case Eligibility	Leukemia	%	Brain/ONS	%	Angio-sarcoma	%	Bladder	%
Eligible	21,861	90.9%	12,394	92.0%	36	100%	47,590	98.4%
Not Eligible*	2,201		1,078		0		795	
Total	24,062		13,472		36		48,385	

***** Cases were not eligible if they were reported by death certificate or autopsy.

Table 2. Geocode Success Rate of Eligilble Cases for Select Cancers

Census Tract Geocode Success	Leukemia	%	Brain/ONS	%	Angio-sarcoma	%	Bladder	%
High Accuracy	20,130	92.1%	11,428	92.2%	36	100%	43,653	91.7%
Zip Code Centroid	1,579	7.2%	918	7.4%	0		3,617	7.6%
Uncodable	152	0.7%	48	0.4%	0		320	0.7%
Total Cases	21,861		12,394		36		47,590	

Table 3. Adjusted* Rate Ratios (RR) for Ambient NATA Benzene Exposure and Leukemia (1979-2002) Geocoded Cases by Full Address.

Variable	RR	95% CI	P- value
Benzene < 1.40 μg/m^3	1.0	-	-
Benzene 1.40 - 2.49 μg/m^3	1.09	1.06 - 1.12	<0.001
Benzene 2.50+ μg/m^3	1.04	0.98 - 1.09	0.214

* Adjusted for benzene category, proportion of the 1990 population below the poverty level, proportion of the 1990 population which is white, sex, and age group.

Table 4a. Adjusted* Rate Ratios (RR) for Ambient NATA Benzene Exposure and Female Leukemia (1979-2002) Geocoded Cases by Full Address.

Variable	RR	95% CI	P- value
Benzene < 1.40 μg/m^3	1.0	-	-
Benzene 1.40 - 2.49 μg/m^3	1.08	1.03 - 1.13	0.001
Benzene 2.50+ μg/m^3	1.11	1.03 - 1.21	0.009

* Adjusted for benzene category, proportion of the 1990 population below the poverty level, proportion of the 1990 population which is white, and age group.

Table 4b. Adjusted* Rate Ratios (RR) for Ambient NATA Benzene Exposure and Male Leukemia (1979-2002) Geocoded Cases by Full Address.

Variable	RR	95% CI	P- value
Benzene < 1.40 μg/m^3	1.0	-	-
Benzene 1.40 - 2.49 μg/m^3	1.10	1.05 - 1.14	<0.001
Benzene 2.50+ μg/m^3	0.97	0.90 - 1.05	0.491

* Adjusted for benzene category, proportion of the 1990 population below the poverty level, proportion of the 1990 population which is white, and age group.

Table 5. Adjusted* Rate Ratios (RR) for Ambient NATA Benzene Exposure and Leukemia (1979-2002) Geocoded Cases by Full Address and Zip Centroid.

Variable	RR	95% CI	P- value
Benzene < 1.40 μg/m^3	1.0	-	-
Benzene 1.40 - 2.49 μg/m^3	1.03	0.99 - 1.06	0.083
Benzene 2.50+ μg/m^3	0.97	0.92 - 1.03	0.298

* Adjusted for benzene category, proportion of the 1990 population below the poverty level, proportion of the 1990 population which is white, sex, and age group.

Table 6a. Adjusted* Rate Ratios (RR) for Ambient NATA Benzene Exposure and Female Leukemia (1979-2002) Geocoded Cases by Full Address and Zip Centroid.

Variable	RR	95% CI	P- value
Benzene < 1.40 μg/m^3	1.0	-	-
Benzene 1.40 - 2.49 μg/m^3	1.03	0.99 - 1.08	0.192
Benzene 2.50+ μg/m^3	1.05	0.97 - 1.13	0.249

* Adjusted for benzene category, proportion of the 1990 population below the poverty level, proportion of the 1990 population which is white, and age group.

Table 6b. Adjusted* Rate Ratios (RR) for Ambient NATA Benzene Exposure and Male Leukemia (1979-2002) Geocoded Cases by Full Address and Zip Centroid.

Variable	RR	95% CI	P- value
Benzene < 1.40 μg/m^3	1.0	-	-
Benzene 1.40 - 2.49 μg/m^3	1.02	0.99 - 1.06	0.229
Benzene 2.50+ μg/m^3	0.91	0.85 - 0.98	0.015

* Adjusted for benzene category, proportion of the 1990 population below the poverty level, proportion of the 1990 population which is white, and age group.

Table 7. Adjusted* Rate Ratios (RR) for Ambient NJDEP Modeled Benzene Exposure and Leukemia (1979-2002) Geocoded Cases by Full Address.

Variable	RR	95% CI	P- value
Benzene < 0.013 μg/m^3	1.0	-	-
Benzene 0.013 - 0.12 μg/m^3	1.07	1.04 - 1.11	<0.001
Benzene $\geq$ 0.13 μg/m^3	1.02	0.88 - 1.18	0.784

* Adjusted for benzene category, proportion of the 1990 population below the poverty level, proportion of the 1990 population which is white, sex, and age group.

Table 8. Adjusted* Rate Ratios (RR) for Ambient NJDEP Modeled Benzene Exposure and Leukemia (1979-2002) Geocoded Cases by Full Address and Zip Centroid.

Variable	RR	95% CI	P- value
Benzene < 0.013 μg/m^3	1.0	-	-
Benzene 0.013 - 0.12 μg/m^3	1.01	0.98 - 1.04	0.399
Benzene ≥ 0.13 μg/m^3	0.97	0.84 - 1.11	0.647

* Adjusted for benzene category, proportion of the 1990 population below the poverty level, proportion of the 1990 population which is white, sex, and age group.

Table 9a. Adjusted* Rate Ratios (RR) for Ambient NATA Vinyl Chloride Exposure and Angiosarcoma (1979-2002).

Variable	RR	95% CI	P- value
Vinyl chloride < 0.01 μg/m^3	1.0	-	-
Vinyl chloride ≥ 0.01 μg/m^3	0.45	0.06 - 3.28	0.429

* Adjusted for vinyl chloride category, proportion of the 1990 population below the poverty level, proportion of the 1990 population which is white, sex, and age group.

Table 9b. Adjusted* Rate Ratios (RR) for Ambient NATA Vinyl Chloride Exposure and Angiosarcoma (1979-2002).

Variable	RR	95% CI	P- value
Vinyl chloride < 0.005 μg/m^3	1.0	-	-
Vinyl chloride ≥ 0.005 μg/m^3	2.35	1.21 - 4.58	0.012

* Adjusted for vinyl chloride category, proportion of the 1990 population below the poverty level, proportion of the 1990 population which is white, sex, and age group.

Table10a. Adjusted* Rate Ratios (RR) for Ambient NJDEP Modeled Vinyl Chloride Exposure and Angiosarcoma (1979-2002).

Variable	RR	95% CI	P- value
Vinyl chloride < 0.01 µg/m^3	1.0	-	-
Vinyl chloride ≥ 0.01 µg/m^3	1.29	0.18 - 9.50	0.800

* Adjusted for vinyl chloride category, proportion of the 1990 population below the poverty level, proportion of the 1990 population which is white, sex, and age group.

Table10b. Adjusted* Rate Ratios (RR) for Ambient NJDEP Modeled Vinyl Chloride Exposure and Angiosarcoma (1979-2002).

Variable	RR	95% CI	P- value
Vinyl chloride < 0.005 µg/m^3	1.0	-	-
Vinyl chloride ≥ 0.005 µg/m^3	1.06	0.37 - 2.99	0.918

* Adjusted for vinyl chloride category, proportion of the 1990 population below the poverty level, proportion of the 1990 population which is white, sex, and age group.

Table 11a. Adjusted* Rate Ratios (RR) for Ambient NATA Vinyl Chloride Exposure and Brain/ONS Cancer (1979-2002) Geocoded Cases by Full Address.

Variable	RR	95% CI	P- value
Vinyl chloride < 0.01 µg/m^3	1.0	-	-
Vinyl chloride ≥ 0.01 µg/m^3	0.92	0.84 - 0.99	0.034

* Adjusted for vinyl chloride category, proportion of the 1990 population below the poverty level, proportion of the 1990 population which is white, and age group.

Table 11b. Adjusted* Rate Ratios (RR) for Ambient NATA Vinyl Chloride Exposure and Brain/ONS Cancer (1979-2002) Geocoded Cases by Full Address.

Variable	RR	95% CI	P- value
Vinyl chloride < 0.005 μg/m^3	1.0	-	-
Vinyl chloride ≥ 0.005 μg/m^3	0.95	0.90 - 0.99	0.012

* Adjusted for vinyl chloride category, proportion of the 1990 population below the poverty level, proportion of the 1990 population which is white, and age group.

Table 12a. Adjusted* Rate Ratios (RR) for Ambient NATA Vinyl Chloride Exposure and Brain/ONS Cancer (1979-2002) Geocoded Cases by Full Address and Zip Centroid.

Variable	RR	95% CI	P- value
Vinyl chloride < 0.01 μg/m^3	1.0	-	-
Vinyl chloride ≥ 0.01 μg/m^3	0.98	0.91 - 1.06	0.611

* Adjusted for vinyl chloride category, proportion of the 1990 population below the poverty level, proportion of the 1990 population which is white, and age group.

Table 12b. Adjusted* Rate Ratios (RR) for Ambient NATA Vinyl Chloride Exposure and Brain/ONS Cancer (1979-2002) Geocoded Cases by Full Address and Zip Centroid.

Variable	RR	95% CI	P- value
Vinyl chloride < 0.005 μg/m^3	1.0	-	-
Vinyl chloride ≥ 0.005 μg/m^3	0.97	0.93 - 1.01	0.136

* Adjusted for vinyl chloride category, proportion of the 1990 population below the poverty level, proportion of the 1990 population which is white, and age group.

Table 13a. Adjusted* Rate Ratios (RR) for Ambient NJDEP Modeled Vinyl Chloride Exposure and Brain/ONS Cancer (1979-2002) Geocoded Cases by Full Address and Zip Centroid.

Variable	RR	95% CI	P- value
Vinyl chloride < 0.01 µg/m^3	1.0	-	-
Vinyl chloride ≥ 0.01 µg/m^3	1.00	0.89 - 1.12	0.973

* Adjusted for vinyl chloride category, proportion of the 1990 population below the poverty level, proportion of the 1990 population which is white, and age group.

Table 13b. Adjusted* Rate Ratios (RR) for Ambient NJDEP Modeled Vinyl Chloride Exposure and Brain/ONS Cancer (1979-2002) Geocoded Cases by Full Address and Zip Centroid.

Variable	RR	95% CI	P- value
Vinyl chloride < 0.005 µg/m^3	1.0	-	-
Vinyl chloride ≥ 0.001 µg/m^3	0.99	0.94 - 1.05	0.763

* Adjusted for vinyl chloride category, proportion of the 1990 population below the poverty level, proportion of the 1990 population which is white, and age group.

Table 14. Adjusted* Rate Ratios (RR) for THM Exposure and Bladder Cancer (1979-2002) Geocoded Cases by Full Address.

Variable	RR	95% CI	P- value
THM = 0 µg/l	1.0	-	-
THM > 0 - 45 µg/l	0.99	0.96 - 1.01	0.413
THM > 45 – 65 µg/l	1.04	1.02 - 1.07	0.001
THM > 6 5 µg/l	1.15	1.11 - 1.18	<0.001

* Adjusted for THM category, proportion of the 1990 population below the poverty level, proportion of the 1990 population which is white, and age group.

Table 15. Adjusted* Rate Ratios (RR) for THM Exposure and Bladder Cancer (1979-2002) Geocoded Cases by Full Address and Zip Centriod.

Variable	RR	95% CI	P- value
THM = 0 µg/l	1.0	-	-
THM > 0 - 45 µg/l	0.96	0.94 - 0.98	0.001
THM > 45 - 65 µg/l	0.98	0.96 - 1.01	0.150
THM > 65 µg/l	1.09	1.06 - 1.12	<0.001

* Adjusted for THM category, proportion of the 1990 population below the poverty level, proportion of the 1990 population which is white, and age group.