# Pollution Sources and Mortality Rates across Rural-Urban Areas in the United States

**Michael Hendryx, PhD; Evan Fedorko; Joel A. Halverson, PhD**

*West Virginia Rural Health Research Center*
*West Virginia University, Morgantown, WV*

***Support for this report was provided by the Office of Rural Health Policy, Health Resources and Services Administration, PHS Grant No. 1 U1CRH10664-01-00.***

**Additional Information**
*Please visit our website: http://wvrhrc.hsc.wvu.edu*

# Pollution Sources and Mortality Rates across Rural-Urban Areas in the United States

**Michael Hendryx, PhD; Evan Fedorko; Joel Halverson, PhD**

*West Virginia Rural Health Research Center*
*West Virginia University, Morgantown, WV*

## Executive Summary

***Background:*** People in rural areas are potentially exposed to a variety of environmental risks from point or non-point pollution sources. A comprehensive assessment of rural environmental pollution sources in relation to health has not yet been undertaken, and the intent of this study is to begin to examine those relationships.

***Methods:*** This study gathered together a set of existing databases to create a new linked database of potential pollution sources and corresponding health outcomes, with particular attention to hazards in rural settings. For most analyses rural was defined as non-metropolitan counties using US Dept. of Agriculture (USDA) rural-urban continuum codes. For some analyses we used those codes to divide counties into metropolitan, micropolitan and non-core counties, and others used the USDA urban-influence codes to define rurality. Data were drawn from several national, public databases as shown in the Box.

***DATA SOURCES***
Environmental Protection Agency (EPA);
Centers for Disease Control and Prevention (CDC);
Department of Energy's Energy Information Administration (EIA);
US Department of Agriculture's (USDA) National Agricultural Statistics Service (NASS);
Multi-Resolution Land Characteristics Consortium's National Land Cover Dataset (NLCD);
Health Resources and Services Administration's Area Resource File (ARF);
Appalachian Regional Commission (ARC);
Healthcare Cost and Utilization Project (HCUP);
U.S. Census Bureau

We created spatial versions of these databases, checked for errors and then utilized spatial software to accurately describe at the county level the types of potential pollution sources present in metropolitan and non-metropolitan areas. Next, we examined the statistical

***Support for this report was provided by the Office of Rural Health Policy, Health Resources and Services Administration, PHS Grant No. 1 U1CRH10664-01-00.***

associations between potential exposures and mortality rates, focusing on all-cause, cardiovascular, respiratory and cancer mortality. These associations were examined through regression modeling to hold constant effects of age, sex, race, and other population characteristics.

***Results:*** Results indicated that rural areas had fewer EPA-monitored pollution sites per county, and fewer fossil fuel plants, relative to urban counties. Nevertheless, rural counties had 65,055 EPA-monitored pollution discharge and monitoring sites. As expected, rural counties had significantly greater exposure to potential agriculture-related pollution. Regression models specific to rural counties indicated that water pollution sources were significantly associated with greater total ($p<.02$) and cancer ($p<.02$) mortality. Rural air pollution monitoring sources were associated with greater cancer mortality rates ($p<.05$). Rural coal mining areas also had higher adjusted mortality rates ($p<.004$ for total mortality and $p<.0002$ for respiratory mortality); these associations reflect population-wide effects, not effects limited to coal miners. Agricultural production was generally associated with lower mortality rates. Among non-metropolitan areas, greater levels of human development (all types of commercial, industrial or residential development excluding parks and agriculture as measured by the Land Cover Dataset) were significantly related to higher adjusted cancer mortality ($p<.003$), which may reflect exposure not only to point pollution sources but also heavier transportation pollution and toxicants related to residential or commercial development . Maps showing pollution sites nationally across rural and urban areas are presented at the end of the report.

***Conclusions:*** The relationships found in this study between pollution sources and mortality rates are statistical associations and do not necessarily reflect causal impacts. However, association between increased levels of pollution and higher mortality risk is not a situation limited to metropolitan areas. Potential pollution sources present in rural areas include air and water pollution from industrial or commercial activity, coal mining, and other types of human development including potentially transportation and residential sources. These activities are highest in non-metropolitan areas that are adjacent to metropolitan areas. Mortality associations in rural areas were strongest for total mortality and cancer mortality. Elevated adjusted mortality rates in non-metropolitan areas in association with a variety of potential pollution sources carries important policy implications regarding the need for effective environmental monitoring and environmental standards. Further research is needed to better understand the types and distributions of pollution in rural areas, and the health consequences that result. Health care professionals who work in rural settings will need to have the appropriate training and resources to diagnose and treat environmentally-instigated or mediated disease. Improved coordination between the health care community and the public health community in rural settings may improve the capacity of rural health care providers to deliver environmentally-sensitive services. More detailed state level maps have been created and are available on the West Virginia Rural Health Research Center website (http://wvrhrc.hsc.wvu.edu).

## INTRODUCTION

Health is influenced by a host of genetic, behavioral, environmental, socioeconomic, and health services variables. Recently, the role of the physical environment has been gaining increased attention as a key threat to public health.(1-5)

These problems are of concern given the limited information about environmental exposures faced specifically by rural populations. Most prior research on environmental health risks has focused on urban populations.(5-6) Rural populations, however, are potentially exposed to a variety of serious environmental risks from point and non-point pollution sources including industrial facilities, animal containment facilities, mining operations, logging and timber activities, petroleum refineries, agricultural activities, incinerators, land fills, sewage treatment facilities, and transportation routes.(7-11) For example, the rural counties of eastern North Carolina have the highest childhood lead poisoning rates in the state.(8) Exposure to agricultural pesticides has adverse impacts on reproductive health and outcomes.(11) Exposure to environmental contaminants increases risks for the most serious forms of illness and premature mortality found in rural populations, including cardiovascular disease, diabetes, and several forms of cancer.(6, 12-17) The total burdens of environmental exposures in rural populations are not known, and carry significant implications for both population health and health care treatment demands that are placed on rural health care systems.

For this study, we focus on pollution sources of a variety of types and their relationship to mortality rates in rural and urban areas. Our intent is, first, to describe the number and type of pollution sources that are present, and second, to test for associations between greater exposures and mortality, holding constant the effects of age, sex, race, and other population characteristics. Outcomes are investigated for all-cause mortality, and for cancer, cardiovascular diseases, and respiratory diseases. We examine associations between potential pollution sources and mortality rates, but recognize that associations do not equate to causality. That is, we cannot conclude that pollution is the cause of any increased mortality rates observed, but rather, that associations only suggest this possibility, which will then require follow-up research to confirm or refute.

## METHODS

**Design:** The study is a retrospective secondary analysis of existing data sources. We employ a natural field experimental design in the comparison of health outcomes as a function of potential environmental exposures in rural and urban settings.

**Rural-urban setting:** Rural-urban setting was measured based on rural-urban continuum codes established by the US Department of Agriculture. For some analyses we used the 9 categories as a scale and for some we divided counties into metropolitan (codes 1-3) and non-metropolitan (codes 4-9) designations. The term "rural" and "non-metropolitan" will be used interchangeably for this study. Regression models with mortality rates as dependent variables and pollution and covariates as independent variables were analyzed separately for non-metropolitan areas. Finally, we also include some findings for the USDA rural definition based on the 12 levels of the urban influence codes (UIC). The Box below summarizes the rural definitions used for various analyses.

***RURAL DEFINITIONS AND ANALYSES***

1. Non-metropolitan counties (codes 4-9) versus metropolitan counties (codes 1-3) from rural-urban continuum codes; used for Tables 2, 5, 6 and 7, and for Figures 1 through 7.
2. Noncore counties (codes 4, 6, 7, 9-12) versus micropolitan counties (codes 3, 5, and 8) versus metropolitan counties (codes 1 and 2) from the urban-influences codes; used for Table 3.
3. Rural-urban continuum code used as 9-point scale; used for Table 4.
4. The 12 urban-influence categories; used for Figures 8 and 9.

**Data sources:** Data are drawn from several public, national datasets. These are described in the Box below. All data were compiled at the county level (N=3,141). The years for each variable are not the same across data sources. Some variables are based on 2000 Census data, some on more recent Census estimates. CDC BRFSS data on smoking rates are from 2003-2006. EPA data are current as of 2008. A summary of the data sources, variables, and years covered by the data is provided in Table 1.

***DATA SOURCES***
Environmental Protection Agency (EPA);
Centers for Disease Control and Prevention (CDC);
Department of Energy's Energy Information Administration (EIA);
US Department of Agriculture's (USDA) National Agricultural Statistics Service (NASS);
Multi-Resolution Land Characteristics Consortium's National Land Cover Dataset (NLCD);
Health Resources and Services Administration's Area Resource File (ARF);
Appalachian Regional Commission (ARC);
Healthcare Cost and Utilization Project (HCUP);
U.S. Census Bureau

**Table 1:** Data sources, variables, years covered

| Data Source | Variables | Years |
|---|---|---|
| EPA TRI | TCL-TRI | 2008 |
| EPA AIRS | TCL-AIRS | 2008 |
| EPA PCS | TCL-PCS | 2008 |
| EPA eGrid | Power plants | 2008 |
| EIA | Coal mining | 1996-2005 |
| USDA | Cropland and animal production | 2002 |
| NLCD | Land use/development | 2001 |
| CDC | Mortality rates | 1997-2005 |
| ARF | Demographic covariates | 2000-2005 |
| ARC | Appalachian county designations | 2006 |
| SID | Hospitalization rates | 2005 |
| BRFSS – CDC | Smoking rates | 2003-2006 |

We compiled data from four EPA program databases: AIRS, PCS, TRI, and eGRID (descriptions of each follow, and are summarized in the next text Box below.)

We gathered information on air pollution releases from the EPA Aerometric Information Retrieval System (AIRS). This is a repository for information about air pollution throughout the United States, and includes reports by various stationary monitoring sources near air pollution such as electric power plants, steel mills, and factories.

The EPA's Permit Compliance System (PCS) contains data on National Pollutant Discharge Elimination System (NPDES) permit holding facilities. The PCS tracks the permit, compliance, and enforcement status of NPDES facilities. Discharges from these facilities occur into water sources.

The EPA Toxics Release Inventory (TRI) contains information about more than 650 toxic chemicals that are used, manufactured, treated, transported, or released into the environment. Chemical manufacturers are required to report the locations and quantities of chemicals stored on-site to state and local governments.

The EPA's eGRID database was the source for information on power plants. The Emissions and Generation Resource Integrated Database (eGRID) is a comprehensive inventory of environmental attributes of electric power systems. It contains information about primary fuel and emissions. The number of power plants per type was recorded for every county. Type of power plants includes fossil fuel plants that have higher emissions (coal, oil and natural gas), as well as nuclear, hydropower, wind and other types.

***EPA SPECIFIC DATA SOURCES***
Aerometric Information Retrieval System (AIRS): air pollution from stationary sources such as factories and power plants;
Permit Compliance System (PCS): waste water discharges from stationary sources;
Toxics Release Inventory (TRI): locations where toxic chemicals are used, manufactured, treated, transported, or released;
Emissions and Generation Resource Integrated Database (eGRID): locations of electrical power plants, grouped into type (coal, oil, natural gas, wind, solar, nuclear, hydropower, geothermal, biofuel).

Modern large scale agriculture relies on pesticides and fertilizers that are known to contain toxicants potentially harmful to human health. Animal facilities are also associated with pollution from waste that may enter ground water or increase levels of ambient particulate matter. We acquired data from the US Department of Agriculture's National Agricultural Statistics Service (NASS), and using SAS 9.1 software, reorganized the data into a format accessible by ArcMap and other spatial software. This dataset contains numerous variables including the percent of each county in acres that was committed to harvested cropland of any type, and the number of animals per acre including cattle (dairy and meat), chickens (broilers and layers), pigs and hogs, and sheep and goats. The dataset also contains information about general farm patterns and farm sizes per county.

The Multi-Resolution Land Characteristics Consortium is a group of federal agencies that work together to fund the production of satellite imagery derived land use/land cover datasets. From this, we obtained the National Land Cover Datasets (NLCD) for the two most recent years in which it was available during the time of data acquisition, 1992 and 2001. Only the 2001 data are analyzed for the current study. Land cover is classified into eight major categories: water or perennial ice and snow; barren land; forest; scrub/shrub; grassland/herbaceous; wetlands; pasture/hay or cultivated crops; and developed land. Developed land is further divided into four levels: open space, or development of low, medium or high intensity. Intensity of development is defined by percent of total cover devoted to impervious surfaces: low (20-49%), medium (50-79%) and high (80% and more).

From the Health Care Utilization Project (HCUP) we obtained the 2001 through 2004 National Inpatient Samples (NIS) for all states participating in the survey. We also examined the contents of the State Inpatient Database (SID) for all states, and obtained the 2005 SID for West Virginia to conduct exploratory tests of whether the SID, which 1) contains almost all hospitalization records and 2) sometimes allows mapping patient residence by zip code, offered statistical improvement over the sample of records available in the NIS that map only the county location of the hospital in detecting associations between hospitalizations and pollution sources.

We attempted to examine other data sources for additional environmental indicators, for example, on timber production, and on other types of mining and drilling in addition to coal mining, but found the data sources were incomplete – some types of indicators were available for some counties but national coverage was not available.

**Dependent variables**: The first dependent variable is average annual all-cause, age adjusted mortality per 100,000 for the years 1997-2005 as provided on the public CDC website. The second dependent variable is average annual age-adjusted cancer mortality per 100,000 for 1997-2005. The third and fourth variables are mortality from any cardiovascular disease, and from any respiratory disease, respectively. Rates are age-adjusted using the standard US Census 2000 population.

For the NIS and SID data we used Major Diagnostic Code (MDC) groups to identify all discharges with 1) a respiratory, 2) cardiovascular, or 3) cancer primary diagnosis. These discharges were expressed on a per capita basis over the entire county population.

**Independent variables:** For each of the EPA datasets (other than eGrid, which was downloaded from it's own website: http://www.epa.gov/cleanenergy/energy-resources/egrid/index.html), we utilized the EPA's Facility Registry System (http://www.epa.gov/enviro/html/fii/ez.html) to download the data in tabular format, state by state. The EPA provides a latitude and longitude coordinate pair for each facility. We plotted these on a map and using Arc Map, the industry standard mapping software published by Environmental Systems Research Institute, we removed extraneous points that plotted outside of the boundaries of their respective state. Errors in location resulted in only a small number of points being removed from the dataset. Once geographic errors were resolved, we compiled the points into a national dataset and utilized Arc Map functionality to count points by type within each US county.

Independent variables drawn from these EPA data included the number of sites per county that are registered with three EPA datasets: 1) the AIRS data on air point pollution sources; 2) the TRI data on toxic-release inventory sites; and 3) the PCS data on water pollution discharge sites. For each of these three sources we calculated the Total Cumulative Load (TCL) as

$$TCL = w_p \Sigma (w_i E_i)$$

Where $E_i$ is scored 0 or 1 to reflect the presence or absence of the environmental pollution source within the county, $w_i$ is a weight equal to the number of such sources, and $w_p$ is a weight equal to the density of the sources as defined by total number of sites divided by the county land area in square miles. A higher TCL score thus reflects greater number of pollution sources relative to county size. A global TCL was found, as were TCLs specific to TRI, AIRS, and PCS data.

We used spatial analysis and the eGrid database to find the number of electricity power plants located in each county by type: coal, wind, hydropower, oil, gas, nuclear, and total. From an air pollution perspective, coal plants emit the highest levels of particulate matter, mercury, sulfur dioxide and other pollutants, followed by oil and then natural gas. Wind, hydro and nuclear plants have no such emissions; they are "clean". We calculated a TCL score specific to coal plants, and one combining fossil fuels (i.e., coal, oil and natural gas.)

We used data from the US Department of Agriculture to measure the percent of the total county acres of harvested land committed to crops of any type, and the number of agricultural animals per acre, including cattle (dairy and meat), chickens (broilers and layers), pigs and hogs, and sheep and goats.

From the EIA we recorded the tons of coal mined at the county level for the years 1996-2005. For regression analyses of coal mining, counties were divided into three categories: those with no coal mining, those with mining below the median tonnage over the combined years, and those with mining above the median value.

Using the NLCD data, we utilized spatial analysis techniques within Arc Map to calculate the percentage of land within each US County taken up by each of the defined land use types. We found the percent of each county that was subject to human development of low, medium, and high intensity, which we then combined into a measure of total percent of developed area. Developed areas excluded agricultural development and developed open land (e.g., parks). Each county, then, has a score representing the percent of its area devoted to developed land of at least low intensity. Differences in the classification scheme between the 1992 and 2001

NLCD products make multiyear direct comparisons difficult; for this study we used only the 2001 NLCD.

**Covariates:** Covariates from the 2005 ARF included the: percent male population, poverty rate, percent high school graduates, percent college graduates, percent African American, percent Native American, percent Hispanic, percent Asian American, and the number of active primary care doctors (MDs and DOs) per 1,000 population. We include both high school and college graduate percents so that the referent variable becomes less than high school; including both high school and college allows for a fuller test of education effects across the range of education levels.

From the ARC we created a regional variable indicating whether the county was in Appalachia. Finally, from CDC BRFSS data, which we supplemented by examining the state public health websites for all 50 states, we created an estimate of each county's adult smoking rate.

We used age and sex of the patient as covariates for the analysis of SID and NIS data, as well as other county-level covariates described above.

**Analysis:** Analysis included descriptive summaries, including EPA pollution sources and power plants, agricultural data, coal mining, and land use cover, for metropolitan and non-metropolitan areas. Data availability and patterns of missing data were examined for SID and NIS hospitalization data. Then, ordinary least squares linear regression models were estimated to examine the association of pollution sources, land use development, and agricultural productivity to age-adjusted mortality rates, controlling for covariates. Models were run for all counties, and separately for non-metropolitan counties.

## RESULTS

A descriptive summary of the potential pollution sources for metropolitan and non-metropolitan areas is provided in Table 2. Metropolitan areas have significantly greater EPA identified point pollution sources, both in terms of raw numbers and TCL scores. They also have significantly greater numbers of power plants of all types except for wind farms. This is not surprising given the greater populations of metropolitan areas. However, rural areas contain thousands of potential pollution sources. There are, for example, 931 fossil-fuel burning power plants in rural areas of the United States. There are 16,574 TRI sites (measuring toxic chemicals), 14,276 AIRS sites (measuring air pollution sites), and 34,214 PSC sites (measuring water pollution sites) in rural areas, a total of 65,055 potential point pollution sites recognized by the EPA.

**Table 2:** Summary of potential pollution sources for metropolitan (N=1090) and non-metropolitan (N=2150) counties.

| | **Non-Metropolitan** | | **Metropolitan** | | **P<** [1] |
|---|---|---|---|---|---|
| | Mean (SD) | Total | Mean (SD) | Total | |
| **TRI sites** | 8.1 (11.7) | 16,574 | 50.2 (111.6) | 54,742 | .0001 |
| **AIRS sites** | 7.0 (14.1) | 14, 267 | 35.1 (129.0) | 38,255 | .0001 |
| **PCS sites** | 16.7 (32.2) | 34,214 | 44.3 (80.2) | 48,284 | .0001 |
| **Coal plants** | .11 (.39) | 99 | .33 (.65) | 355 | .0001 |

| | | | | | |
|---|---|---|---|---|---|
| **Oil plants** | .22 (.62) | 444 | .32 (.94) | 353 | .0001 |
| **Natural gas plants** | .19 (.51) | 388 | .92 (2.5) | 1,008 | .0001 |
| **Fossil fuel power plants** [2] | .52 (.93) | 931 | 1.57 (3.0) | 1,716 | .0001 |
| **Hydropower plants** | .34 (1.34) | 702 | .59 (1.87) | 641 | .0001 |
| **Wind plants** | .06 (.65) | 121 | .09 (1.3) | 94 | ns |
| **Nuclear plants** | .007 (.08) | 14 | .04 (.21) | 46 | .0001 |
| **Total power plants** [3] | 1.01 (1.91) | 2,061 | 2.67 (5.1) | 2,913 | .0001 |
| **TCL – Total** | .065 (.23) | -- | .34 (1.07) | -- | .0001 |
| **TCL – TRI** | .018 (.08) | -- | .11 (.26) | -- | .0001 |
| **TCL – AIRS** | .019 (.14) | -- | .14 (.90) | -- | .0001 |
| **TCL – PCS** | .029 (.06) | -- | .08 (.14) | -- | .0001 |
| **TCL – Fossil fuel power plants** | .0008 (.002) | -- | .0035 (.02) | -- | .0001 |
| **Coal mining counties** | -- | 99 | -- | 40 | na |
| **Percent developed land** | 1.43 | -- | 8.63 | -- | .0001 |
| **Percent acres in cropland** [4] | 23.3 | -- | 19.4 | -- | .0001 |
| **Animals per acre** [5] | 8.0 (0.7) | -- | 7.9 (0.9) | | ns |

1. Non-metropolitan and metropolitan means significantly different.
2. Fossil fuel power plants is the sum of coal, oil, and natural gas plants.
3. Total power plants include those listed plus geothermal, solar, biofuel and other.
4. Crops include corn, barley, wheat, hay, beans, cotton, tobacco, and multiple other types of grains, fruits and vegetables.
5. Animals include beef and dairy cows, pigs and hogs, chickens (broilers and layers), sheep and goats.

Acres committed to cropland were significantly higher in rural areas, but number of animals per acre was not, which may reflect greater relative density of agricultural animals in metropolitan counties which are on average smaller in size than rural counties.

Table 3 repeats the summary of pollution sources found in Table 2, but breaks down non-metropolitan counties into micropolitan and noncore counties. The pattern of results found in Table 2 is largely repeated in Table 3: metropolitan sites have greater average number and concentration of pollution sites per county than micropolitan areas, which in turn have higher number and concentration per county than noncore areas. Exceptions to this pattern are that micropolitan areas have the highest percent of land committed to agriculture, and the highest concentration of agricultural animals per acre.

**Table 3:** Summary of potential pollution sources for metropolitan (N=1,090), micropolitan (675), and non-core (N=1,376) counties.

| | Non-Core | | Micropolitan | | Metropolitan | | P< [5] |
|---|---|---|---|---|---|---|---|
| | Mean (SD) | Total | Mean (SD) | Total | Mean (SD) | Total | |
| TRI sites | 4.2 (5.7) | 5,811 | 15.9 (16.0) | 10,763 | 50.2 (111.6) | 54,742 | .0001 |
| AIRS sites | 5.2 (11.4) | 7,182 | 10.5 (17.8) | 7,085 | 35.1 (129.0) | 38,255 | .0001 |
| PCS sites | 12.8 (29.0) | 17,665 | 24.5 (36.8) | 16,549 | 44.3 (80.2) | 48,284 | .0001 |
| Coal plants | .07 (.30) | 96 | .20 (.52) | 137 | .33 (.65) | 355 | .0001 |
| Oil plants | .20 (.62) | 281 | .24 (.63) | 163 | .32 (.94) | 353 | .0004 |
| Natural gas plants | .13 (.42) | 182 | .31 (.65) | 206 | .92 (2.5) | 1,008 | .0001 |
| Fossil fuel power plants [1] | .41 (.83) | 559 | .75 (1.1) | 506 | 1.57 (3.0) | 1,716 | .0001 |
| Hydropower plants | .23 (.85) | 322 | .56 (2.0) | 380 | .59 (1.87) | 641 | .0001 |
| Wind plants | .06 (.76) | 85 | .05 (.34) | 36 | .09 (1.3) | 94 | .71 |
| Nuclear plants | .005 (.07) | 7 | .01 (.10) | 7 | .04 (.21) | 46 | .0001 |
| Total power plants [2] | .76 (1.5) | 1,051 | 1.5 (2.5) | 1,010 | 2.67 (5.1) | 2,913 | .0001 |
| TCL – TRI | .01 (.07) | -- | .03 (.09) | -- | .11 (.26) | -- | .0001 |
| TCL – AIRS | .02 (.16) | -- | .02 (.10) | -- | .14 (.90) | -- | .0001 |
| TCL – PCS | .02 (.06) | -- | .04 (.05) | -- | .08 (.14) | -- | .0001 |
| TCL – Fossil fuel power plants | .0008 (.002) | -- | .001 (.003) | -- | .0035 (.02) | -- | .0001 |
| Coal mining counties | -- | 67 | -- | 32 | -- | 40 | .32 |
| Percent developed land | 1.1 (2.6) | -- | 2.2 (2.9) | -- | 8.6 (13.5) | -- | .0001 |
| Percent acres in cropland [3] | 22.2 (23.8) | -- | 25.5 (25.0) | -- | 19.4 (21.0) | -- | .0001 |
| Animals per acre [4] | 7.6 (32.0) | -- | 9.0 (33.3) | -- | 7.9 (30.8) | -- | .64 |

1. Fossil fuel power plants is the sum of coal, oil, and natural gas plants.
2. Total power plants include those listed plus geothermal, solar, biofuel and other.
3. Crops include corn, barley, wheat, hay, beans, cotton, tobacco, and multiple other types of grains, fruits and vegetables.

4. Animals include beef and dairy cows, pigs and hogs, chickens (broilers and layers), sheep and goats.
5. Significant difference among the three counties groups.

National maps showing the distribution of point pollution densities, coal mining intensity, agricultural development, and land use development are shown in Figures 1 through 7. These maps use the rural-urban continuum codes to identify metropolitan and non-metropolitan counties. Figures 1 through 3 show that toxic chemical use, air pollution sites, and water pollution sites, are present at higher TCL levels in the eastern half of the country.

More specifically, the maps show that toxic chemical use is highest in areas of central Appalachia up through the New York City and Boston areas, along Lake Michigan and Lake Ontario, areas of Georgia and Indiana, and major metropolitan areas including St. Louis, San Francisco, and others (Figure 1.) Air pollution sites are higher in the area stretching from central Appalachia through Washington DC (Figure 2). Water pollution sites are higher in areas of southern Louisiana, north Alabama, and in an area stretching from Ohio and North Carolina up to New York. Coal mining is concentrated in central and northern Appalachia, and in selected locations in the Midwest and West (Figure 4).

The maps further show that agriculture production is concentrated in the Midwest, along with selected other areas such as the Palouse of the Northwest, the central valley in California, along the Mississippi River, and other places (Figure 5). Livestock production is concentrated in Iowa, Kansas, other Midwest areas, as well as places in Kentucky and North Carolina (Figure 6.) Finally, poultry production is concentrated in the South in areas extending from northeast Texas through Maryland (Figure 7).

**Figure 1:** Toxic Release Inventory Facilities per square mile.

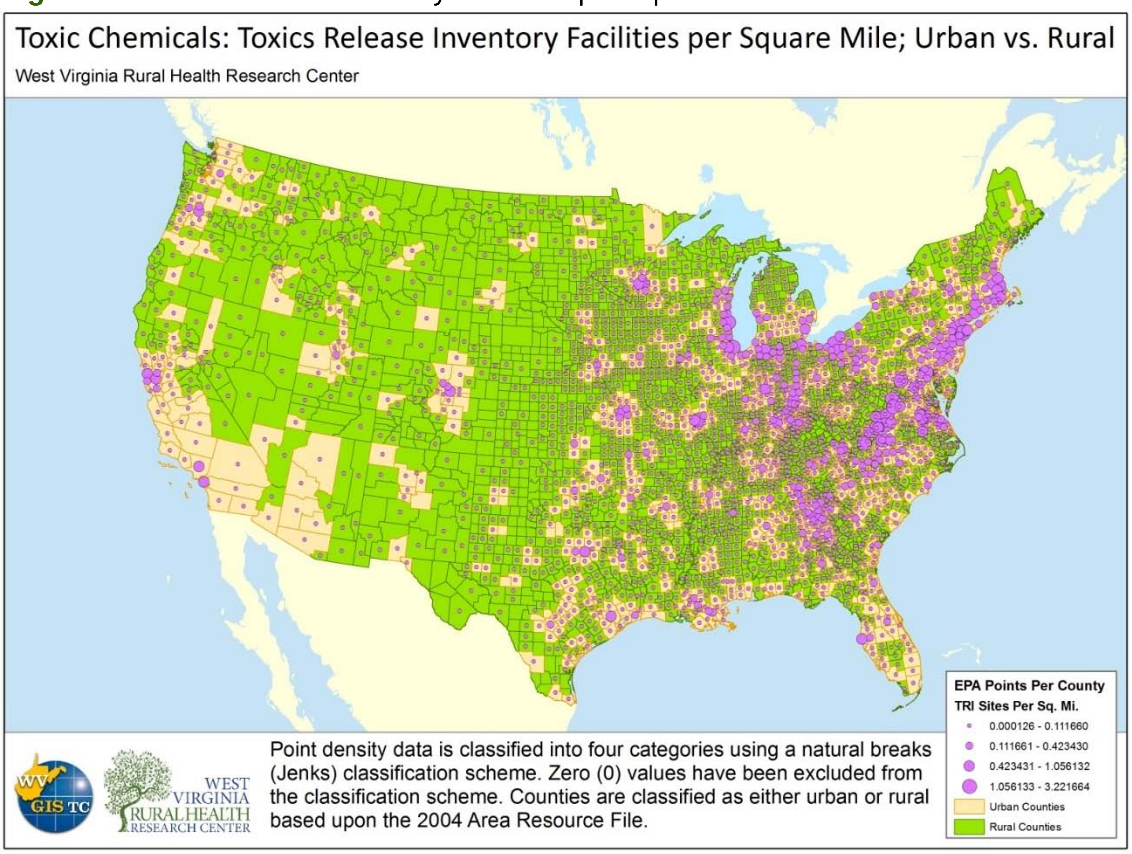


**Figure 2:** Aerometric Information Retrieval System (AIRS) sites per square mile.

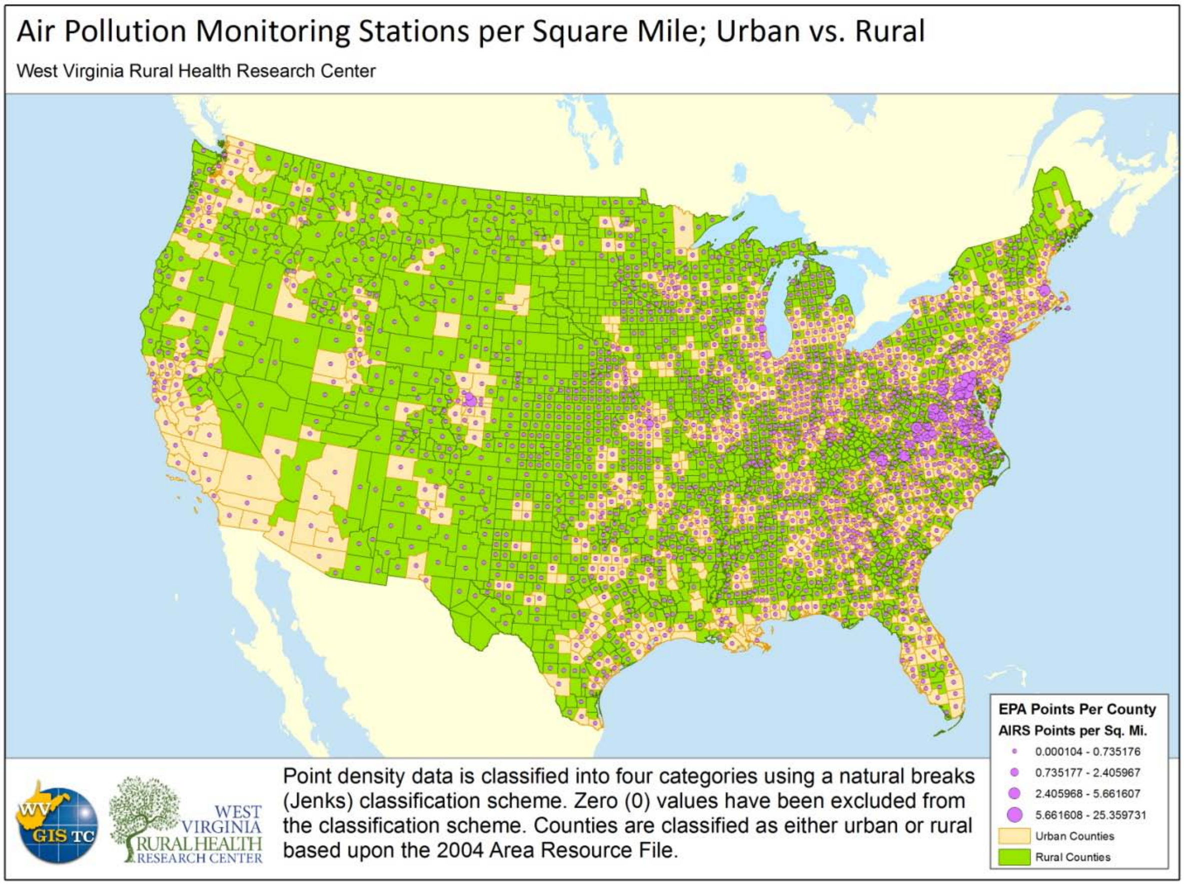



**Figure 3:** Permit Compliance System (PCS) sites per square mile.

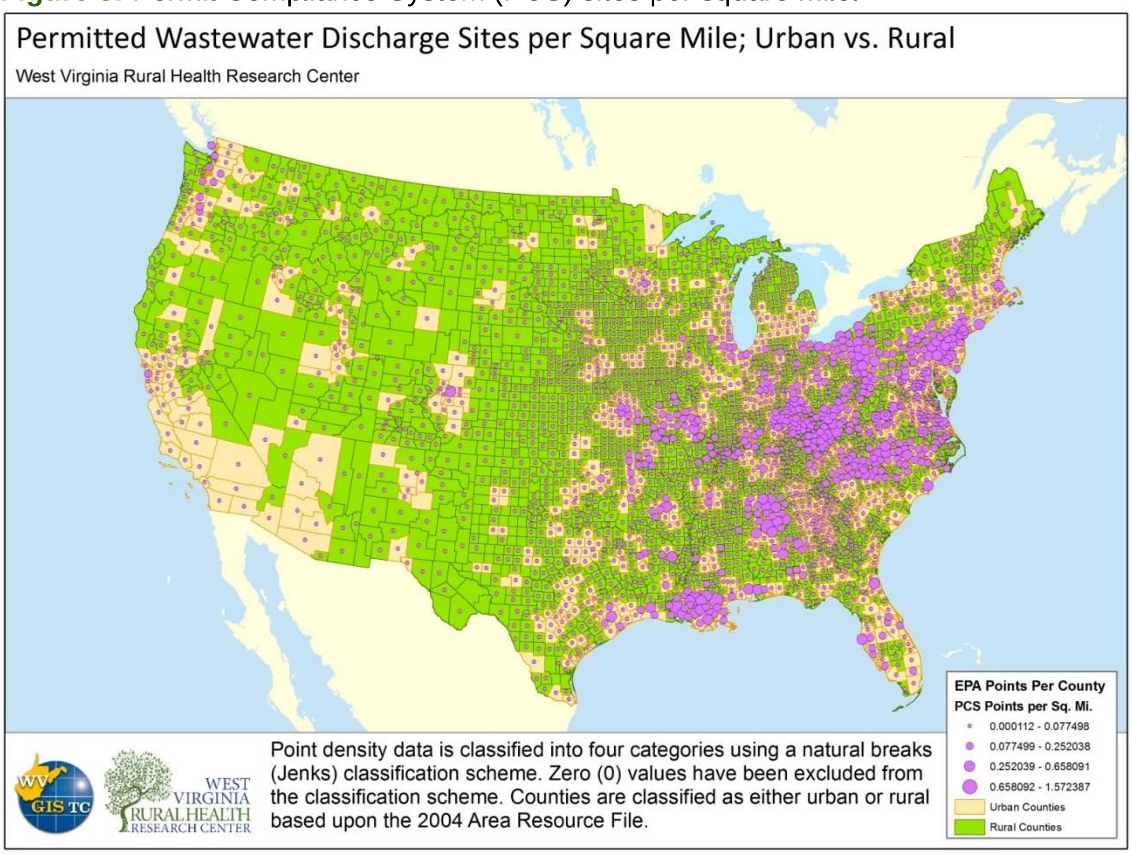


**Figure 4:** Coal production by county.

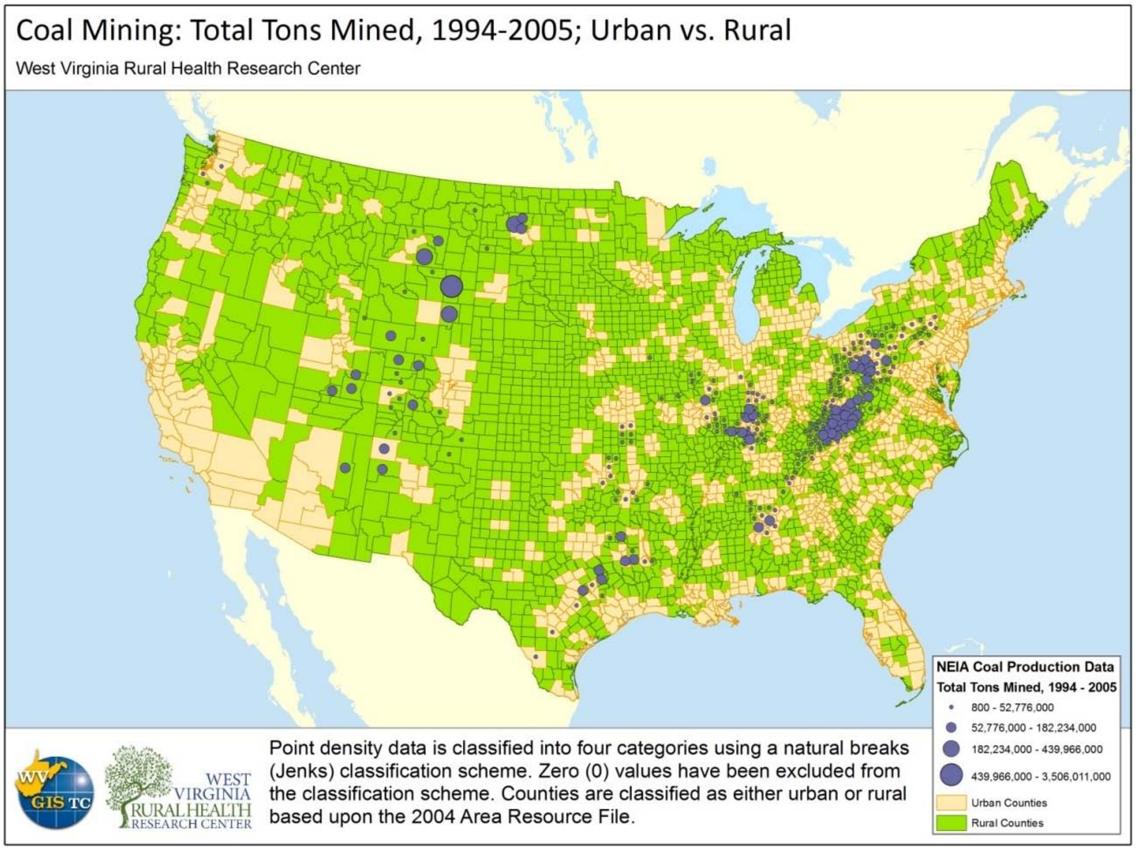



**Figure 5:** Percentage of county acreage in cultivated croplands.

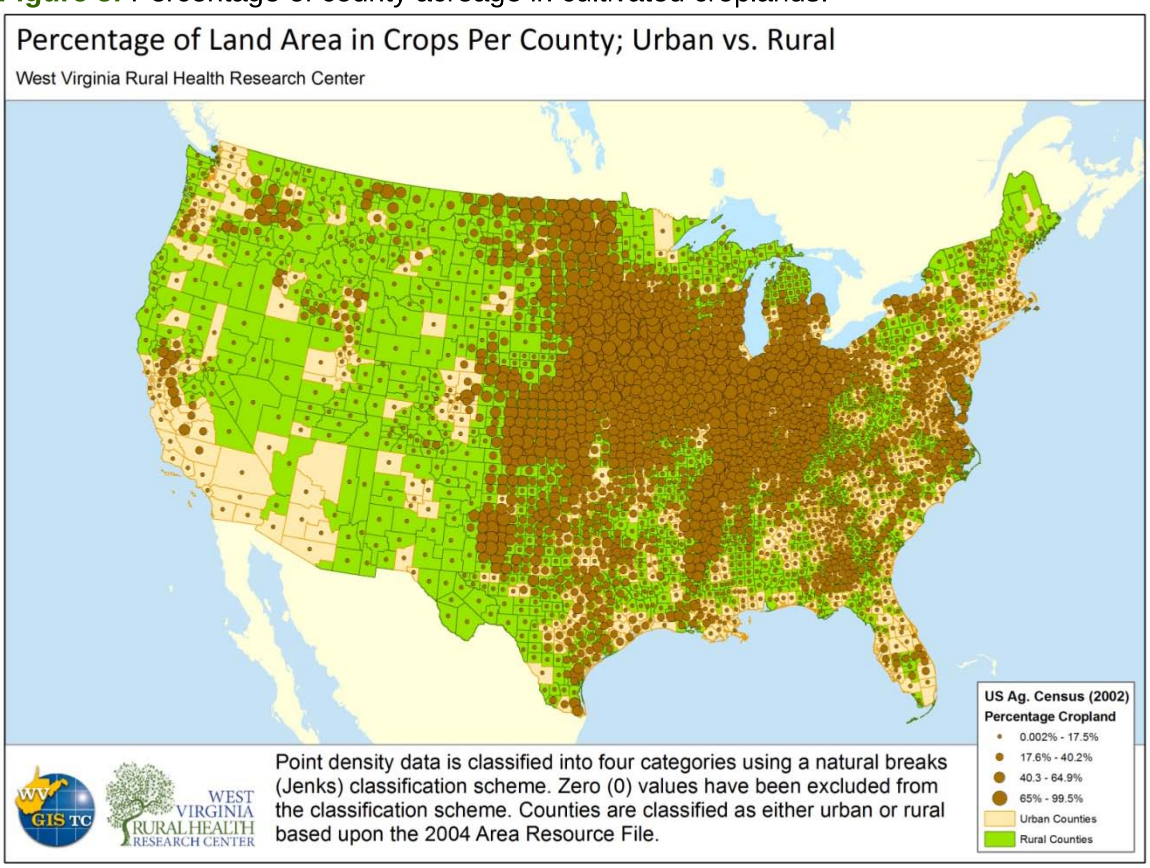


**Figure 6:** Number of livestock (cattle, sheep and pigs) per square mile.

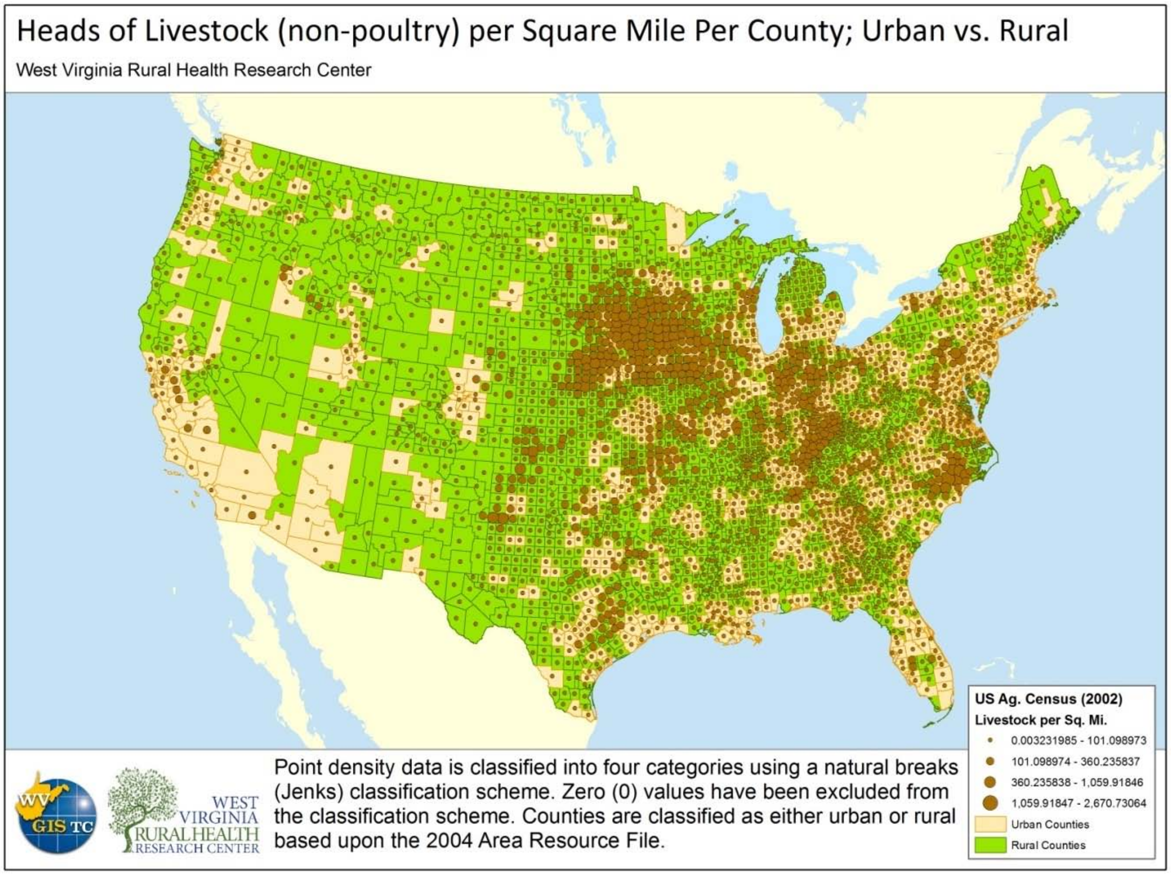
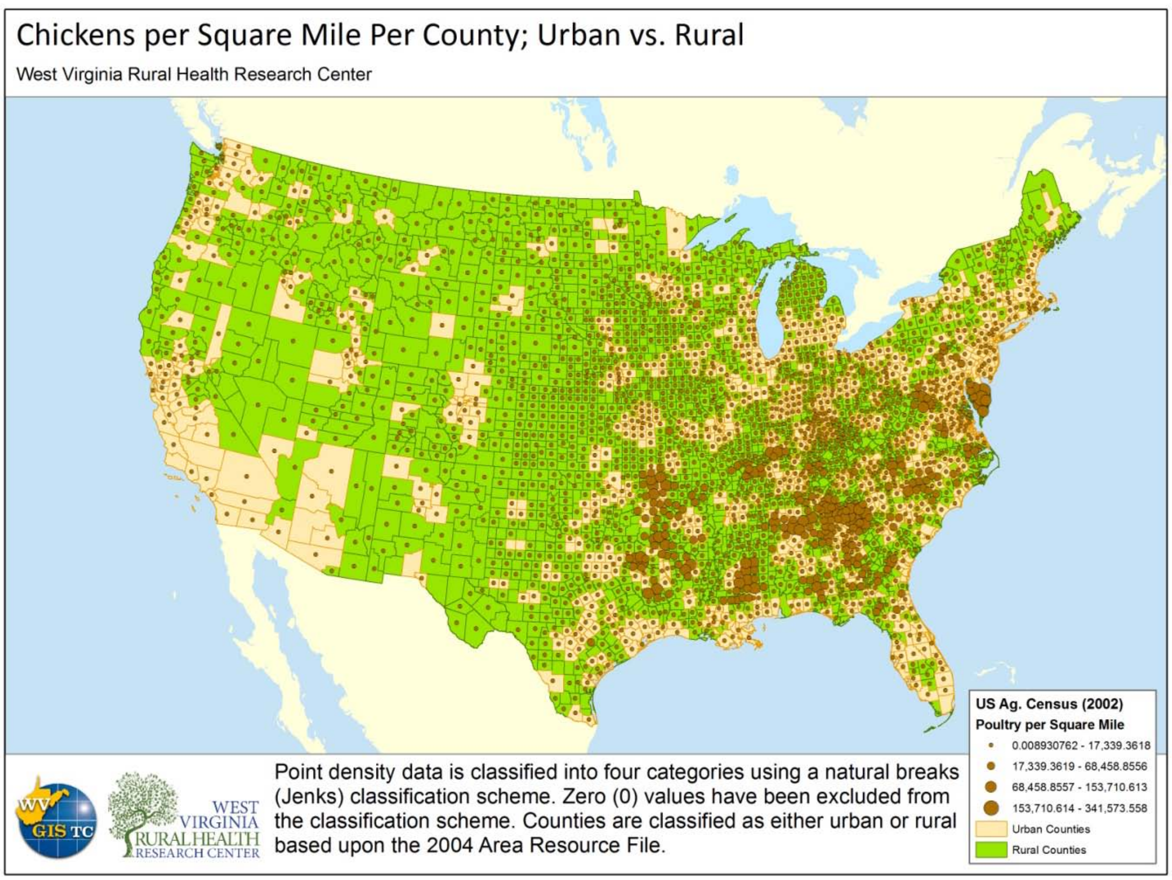



**Figure 7:** Number of poultry per square mile.



***Industrial pollution*:** Because agricultural sources and industrial sources of pollution are different, we constructed separate regression models for the association of these sources to mortality rates. Preliminary regression models for industrial pollution sources included the presence of EPA sites and fossil-fuel power plants (coal, oil and natural gas) as independent variables. When all variables were included, the EPA sites were significantly related to mortality indicators but the power plants did not add independent variance. Power plants and AIRS sites to some extent overlap. Therefore, we excluded power plants from the final models.

In results not shown here, we found that the total TCL score was significantly related to mortality rates for total, cancer, and respiratory mortality, for both all counties and non-metropolitan counties. An overall TCL effect was not found for cardiovascular disease mortality. We then broke down the TCL score into its three component parts (AIRS, TRI and PCS) to determine which were associated with higher adjusted mortality. Results of these final regression models are summarized in Tables 4 and 5. These results present statistical associations only, and do not provide definitive evidence that pollution sources are the cause of observed mortality rates; such a causal impact is possible but requires additional research to confirm or refute.

Table 4 shows the analysis for non-metropolitan counties only. Controlling for covariates, greater concentration of air (AIRS) and water (PCS) pollution sites were independently related to higher adjusted cancer mortality rates, and water pollution sites was related to higher total mortality rates. Non-metropolitan counties with high coal mining had higher total and respiratory mortality.

**Table 4:** Regression results for *non-metropolitan* counties (N=2,051).* Cells are unstandardized regression coefficients with standard errors in parentheses. Independent variables of interest are industrial pollution sources.

| **Mortality:** | **Total** | | **Cancer** | | **Cardiovascular** | | **Respiratory** | |
|---|---|---|---|---|---|---|---|---|
| ***Model 1*** | Estimate (SE) | P< | Estimate (SE) | P< | Estimate (SE) | P< | Estimate (SE) | P< |
| **TCL- AIRS** | 6.01 (19.00) | .75 | 12.56 (6.38) | .05 | 2.68 (12.57) | .83 | 6.29 (5.54) | .26 |
| **TCL – PCS** | 76.31 (32.14) | .02 | 26.43 (10.79) | .02 | 13.35 (21.27) | .53 | 7.39 (9.37) | .43 |
| **TCL – TRI** | 17.20 (34.60) | .62 | -22.90 (11.62) | .05 | -12.29 (22.90) | .59 | -1.70 (10.09) | .87 |
| ***Model 2*** | | | | | | | | |
| **Low coal mining** | 10.64 (8.63) | .22 | 5.17 (2.91) | .08 | .81 (5.72) | .89 | 1.24 (2.51) | .62 |
| **High coal mining** | 22.84 (7.95) | .004 | 3.52 (2.67) | .19 | 1.17 (5.26) | .82 | 8.76 (2.31) | .0002 |

*Covariates included are: smoking rate, percent male population, primary care physicians per 1,000 population, poverty rate, rural-urban continuum code, percent race/ethnicity groups (African American, Native American, Asian American, Hispanic), high school education rate, college education rate, Appalachian county.

Table 5 repeats the analysis for all counties. Controlling for covariates, greater AIRS density was associated with higher total mortality rates, and the PCS density variable was related to higher total and cancer mortality. Table 5 also shows the results for the national coal mining analysis. Areas of heavier coal mining were associated with higher total, cancer and respiratory mortality.

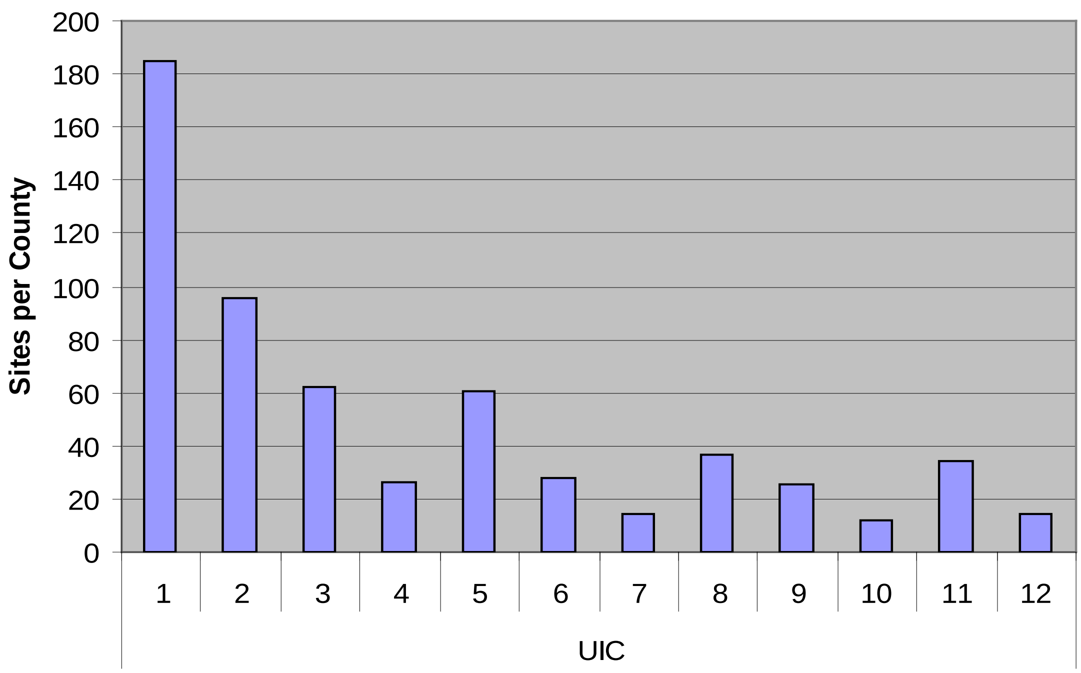
**Table 5:** Regression results for *all* counties (N=3,141).* Cells are unstandardized regression coefficients with standard errors in parentheses. Independent variables of interest are industrial pollution sources.

| **Mortality:** | **Total** | | **Cancer** | | **Cardiovascular** | | **Respiratory** | |
|---|---|---|---|---|---|---|---|---|
| ***Model 1*** | Estimate (SE) | P< | Estimate (SE) | P< | Estimate (SE) | P< | Estimate (SE) | P< |
| **TCL- AIRS** | 6.51 (2.52) | .01 | 1.46 (.83) | .08 | -.19 (1.65) | .91 | .95 (.71) | .19 |
| **TCL – PCS** | 51.06 (14.53) | .0004 | 16.02 (4.80) | .0008 | 10.36 (9.55) | .28 | -2.75 (4.10) | .50 |
| **TCL – TRI** | 13.72 (9.00) | .13 | -4.04 (2.97) | .18 | -5.85 (5.91) | .33 | .94 (2.54) | .71 |
| ***Model 2*** | | | | | | | | |
| **Low coal mining** | 6.97 (6.56) | .29 | 3.70 (2.16) | .09 | 2.39 (4.29) | .58 | -.29 (1.84) | .87 |
| **High coal mining** | 20.16 (6.38) | .002 | 4.20 (2.10) | .05 | 4.79 (4.17) | .25 | 6.31 (1.79) | .0004 |

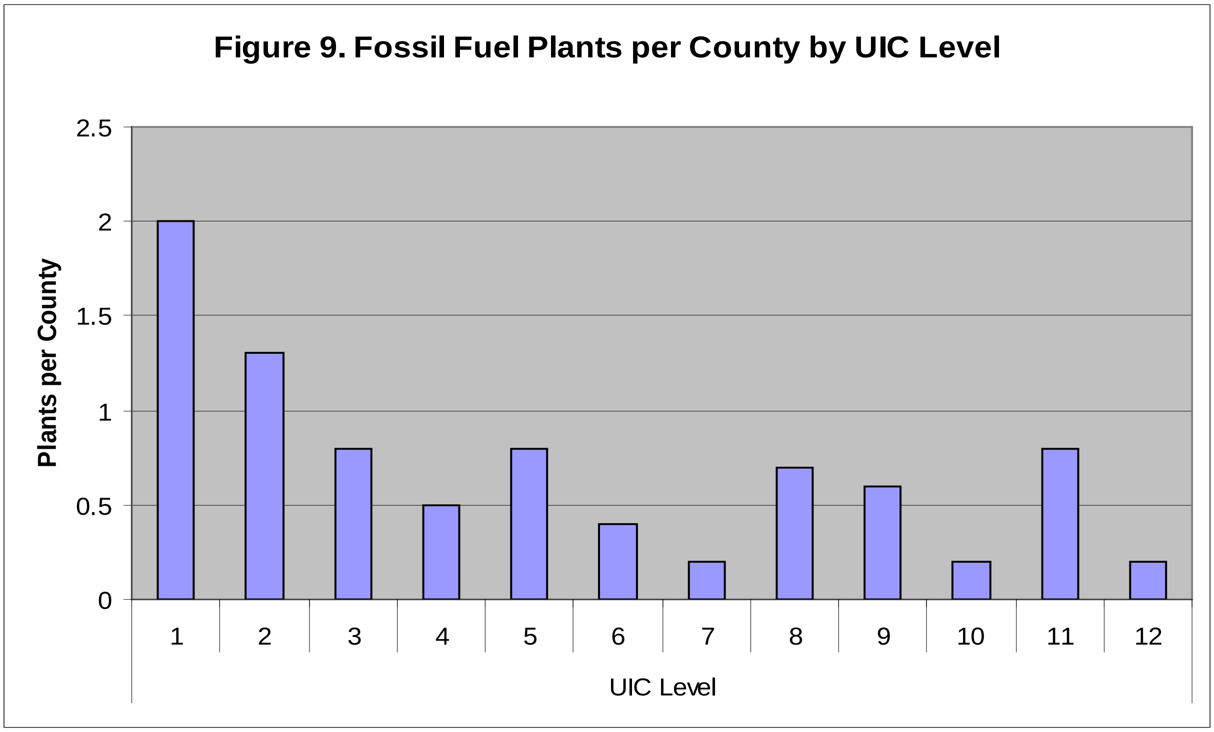
Concerning other covariates, we found effects on mortality in ways consistent with prior research. Controlling for other variables, greater mortality risk was present in rural areas in counties characterized by higher percentages of African American or Native American residents, higher poverty rates, lower education rates, and higher smoking rates. The full regression results with all covariates for all models are not shown here to conserve space.

Figures 8 and 9 show the average number of EPA sites and the average number of fossil fuel plants per county, respectively, along the urban-rural continuum. Urban Influence Codes 1 through 12 are used for these figures. The EPA site variable is the sum of the TRI, AIRS and PCS sites per county. Fossil fuel plants are the sum of coal, oil and natural gas power plants per county. For both Figures, the two metropolitan county groups (codes 1 and 2) have the greatest number per county. The next highest numbers are found for the micropolitan areas (codes 3, 5 and 8), and the smallest numbers for the non-core areas. Note, however, that counties with UIC code 11 (a non-core, non-adjacent area with a town of at least 2,500) contain power plants at levels equal to the larger micropolitan areas.

**Figure 8. EPA sites per County by UIC Level**




Urban Influence Codes (UIC)
1. Large metro area of 1+ million residents
2. Small metro area of less than 1 million residents
3. Micropolitan area adjacent to large metro area
4. Noncore adjacent to large metro area
5. Micropolitan area adjacent to small metro area
6. Noncore adjacent to small metro area and contains a town of at least 2,500 residents
7. Noncore adjacent to small metro area and does not contain a town of at least 2,500 residents
8. Micropolitan area not adjacent to a metro area
9. Noncore adjacent to micro area and contains a town of at least 2,500 residents
10. Noncore adjacent to micro area and does not contain a town of at least 2,500 residents
11. Noncore not adjacent to metro or micro area and contains a town of at least 2,500 residents
12. Noncore not adjacent to metro or micro area and does not contain a town of at least 2,500 residents.




Urban Influence Codes (UIC)
1. Large metro area of 1+ million residents
2. Small metro area of less than 1 million residents
3. Micropolitan area adjacent to large metro area
4. Noncore adjacent to large metro area
5. Micropolitan area adjacent to small metro area
6. Noncore adjacent to small metro area and contains a town of at least 2,500 residents
7. Noncore adjacent to small metro area and does not contain a town of at least 2,500 residents
8. Micropolitan area not adjacent to a metro area
9. Noncore adjacent to micro area and contains a town of at least 2,500 residents
10. Noncore adjacent to micro area and does not contain a town of at least 2,500 residents
11. Noncore not adjacent to metro or micro area and contains a town of at least 2,500 residents
12. Noncore not adjacent to metro or micro area and does not contain a town of at least 2,500 residents.

***Agricultural production*:** The regression models for agricultural production in non-metropolitan counties are summarized in Table 6. We found that greater crop production was associated with lower total, cancer, and respiratory mortality rates. Higher numbers of agricultural animals per acre were not related to mortality rates.

**Table 6:** Regression results for non-metropolitan counties, models containing agricultural production data.*

| Mortality: | Total | | Cancer | | Cardiovascular | | Respiratory | |
|---|---|---|---|---|---|---|---|---|
| | Estimate (SE) | P< | Estimate (SE) | P< | Estimate (SE) | P< | Estimate (SE) | P< |
| **Animals per acre** | .05 (.06) | .40 | -.02 (.02) | .22 | .06 (.03) | .07 | .01 (.01) | .54 |
| **Percent acres in crops** | -73.71 (9.21) | .0001 | -7.07 (2.51) | .005 | -8.79 (4.83) | .07 | -12.88 (2.16) | .0001 |

*Covariates included are: smoking rate, percent male population, primary care physicians per 1,000 population, poverty rate, rural-urban continuum code, percent race/ethnicity groups (African American, Native American, Asian American, Hispanic), high school education rate, college education rate, Appalachian county.

***Land development*:** Table 7 presents the regression results when the key independent variable was the percent of developed land per county, among counties in non-metropolitan areas. Developed land includes the percent of the area devoted to human development as measured by at least 20% impervious surface land cover. Counties with greater development had higher cancer mortality rates, adjusting for the same set of covariates as in other models, including the rural-urban continuum code.

**Table 7:** Regression results for non-metropolitan counties, analysis of developed land percent from the NLCD.*

| Mortality: | Total | | Cancer | | Cardiovascular | | Respiratory | |
|---|---|---|---|---|---|---|---|---|
| | Estimate (SE) | P< | Estimate (SE) | P< | Estimate (SE) | P< | Estimate (SE) | P< |
| **Percent developed land** | 120.6 (77.0) | .12 | 62.45 (20.63) | .003 | 13.18 (39.78) | .74 | 2.26 (17.90) | .90 |

*Covariates included are: smoking rate, percent male population, primary care physicians per 1,000 population, poverty rate, rural-urban continuum code, percent race/ethnicity groups (African American, Native American, Asian American, Hispanic), high school education rate, college education rate, Appalachian county.

***Hospitalization data*:** Our analysis of the SID and NIS data led us to conclude that the data quality and structure were not adequate to include hospitalizations as a dependent variable in analyses. The NIS data document only the county location of the hospital and not the residence of the patient; clearly, many patients cross county lines to receive inpatient hospital care, especially for more serious or complicated cases. The SID data are available for the greatest number of states in 2005 (n=24 states); our examination of the state-specific SID data for 2005 revealed that only two states, West Virginia and Arkansas, populated the patient zip code data field. As a test, we conducted an analysis of the West Virginia SID and discovered that 85% of the in-state zip code values seemed to be in error; they were not recognized by the US Postal Service or other zip code list web sites as valid. We decided not to pursue further analysis of hospitalization data.

## DISCUSSION

Although rural areas contain fewer EPA sites and power plants per county or as a function of county size compared to urban areas, they nevertheless contain thousands of these potential pollution sources. In regression models specific to non-metropolitan counties, a greater exposure to potential sources of anthropogenic air and/or water pollution was associated with higher total and cancer mortality rates, adjusting for age, race, smoking, poverty, education and other covariates. Significant association between pollution sources and higher mortality is a phenomenon present in both urban and rural settings.

The presence of coal mining as an independent risk factor for total, cancer, and respiratory mortality highlights the potential environmental threats posed by the extraction, processing and transportation of coal. Although the number of coal fired power plants was not an independent predictor in the mortality models, the AIRS dataset includes power plants in addition to other industrial air pollution sources and so captures the effects of this pollution source.

Counties with higher levels of development were also associated with higher cancer mortality rates. Furthermore, this effect was found when the analysis was limited to non-metropolitan counties and when rural-urban continuum codes were included as a covariate. This finding indicates that the rural-urban coding does not capture all health relevant variability in the rural-urban environment, and introduces a new environmental variable in the understanding of public health outcomes: land development. The association of greater development with higher cancer mortality in non-metropolitan areas suggests that environmental health concerns associated with human development, such as traffic congestion or industry, are not limited to urban areas.

Regarding agriculturally-related effects, the presence of greater cropland production was related to lower mortality for most investigated types after adjusting for covariates. There may be features of the agricultural lifestyle (e.g., outdoor physical activity, lower risk behaviors such as alcohol consumption), environmental benefits in agricultural landscapes (e.g., cleaner air) or socioeconomic correlates of an agricultural population that lead to this protective effect.(18,19) When it comes to age-adjusted mortality considered across the entire population, the rural agricultural environment is a healthy environment. Within agricultural areas, however, it is still important in future research to examine possible health problems that exist for some populations (e.g., Latino migrant workers, communities with well water impacted by pesticide runoff) or some geographic areas. Pesticide exposures for out-of-county migrant workers is a well established risk and should not be overlooked in considering the results of our study.

This study provides only an overview of potential relationships between pollution sources as captured from the EPA, the NLCD, the EIA and the USDA, and a limited set of health outcomes. We counted only the number of pollution sources per square mile, and no detailed environmental quality data were analyzed. In addition, associations were examined only at the county level. The associations identified here suggest, but do not prove, causal relationships between the pollution sources measured and mortality. There is great variability in the types of chemicals, facilities, and discharge quantities from site to site. Some sites are no doubt benign. Others may be extremely hazardous. There is much future research that could be done. Future studies may examine the variability in types of sites; for example, research may select only TRI sites with documented releases, or PCS releases of certain types or quantities. Additional research can also be undertaken on other pollution sources not included in this report (e.g., uranium mining) and other health outcomes (e.g., hospitalizations, birth outcomes such as low birth weight, self-reported health). Greater refinements in coding of mortality (e.g., investigations into forms of cancer or of chronic or acute circulatory or respiratory disease) may be illuminating. For example, previous research has found a significant population coal mining exposure effect specifically for lung cancer.(20) Finally, geographic studies may be undertaken to understand the spatial distribution of environmental risks for identification of extreme exposure locations (i.e., “hot spots”) in rural areas. Further spatial refinements of TCL scores may also be undertaken to take into account, in calculating a given score, the scores of surrounding counties; this can be done using established spatial correlation techniques and allow for the visual analysis of areas where certain outcomes (e.g., cancer mortality) or certain associations (e.g., pollution and mortality) are clustered as well as areas where there are spatial outliers that suggest areas in need of further investigation.

Variability in years covered across databases is another source of error in the measures. To some extent we assume that both environmental threats and population health parameters are stable over time – areas with high mortality one year will tend to be high in other years, or a pollution source present in one year will be present in others. We also do not account for changes over time in population migration patterns, that people may face exposures in one geographic location and manifest illness only later in another location. The imprecision in the temporal and spatial distribution of pollution and disease could be expected, in general, to weaken associations that really exist, and so the effects that we have found may be underestimated, and others not detected may be issues of poor statistical precision. Nevertheless, future research should endeavor to specify temporal and spatial relationships more precisely, for example, by linking environmental exposures from previous years to health outcomes in later years under conditions where environmental exposure is expected to have delayed rather than immediate effects on the health outcomes of interest.

*Policy implications*

From a policy perspective, the results of the study offer a beginning point to examine both environmental and health care policies that can be enacted or modified to improve population health based on lowering environmental risks. For example, the finding of greater mortality related to coal mining activity suggests the need for improved environmental monitoring around mining sites. In West Virginia, the EPA maintains 22 air monitoring stations around the state, but none are located in communities that are primarily coal mining localities. Additional monitors may be indicated for these settings. There may also be a need for improved water quality testing for private or public water systems in rural areas.

Modifications not only to environmental policy but to health care policy may also be indicated as this line of research goes forward. The epidemiology of disease may be expected to change as levels of environmental pollution increase, and this will have corresponding implications for the types of medical professionals that will be in greater demand. Health care professionals who work in rural settings will need to have the appropriate training and resources to diagnose and treat environmentally-instigated or mediated disease. In this capacity, it will be important for providers to understand the potential environmental threats that exist in their geographic practice area, so that they may make informed diagnostic and treatment decisions. Improved coordination between the health care community and the public health community in rural settings may improve the capacity of rural health care providers to deliver environmentally-sensitive services.

## REFERENCES

1. Pope CA, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, et al. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. JAMA 2002;287:1132-1141.
2. Schulz A, Northridge ME. Social determinants of health: implications for environmental health promotion. Health Education and Behavior 2004;31:455-471.
3. Duderstadt KG. Environmental health policy and children's health. Journal of Pediatric Health Care 2006;20:411-413.
4. Thornton JW, McCally M, Houlihan J. Biomonitoring of industrial pollutants: health and policy implications of the chemical body burden. Public Health Reports 2002;117:315-323.
5. Northridge ME, Stover GN, Rosenthal JE, Sherard D. Environmental equity and health: understanding complexity and moving forward. American Journal of Public Health 2003;93:209-214.
6. Wellenius GA, Schwartz J, Mittleman MA. Particulate air pollution and hospital admissions for congestive heart failure in seven United States cities. American Journal of Cardiology 2006;97:404-408.
7. Hendryx M, Ahern M. Relations between health indicators and residential proximity to coal mining in West Virginia. American Journal of Public Health 2008;98:669-671.
8. Hanchette CL. The political ecology of lead poisoning in eastern North Carolina. Health and Place 2008;14:209-216.
9. Khudar S, Milz SA, Bisesi M, Vincent R, McNulty W, Czajkowski K. Health survey of residents living near farm fields permitted to receive biosolids. Archives of Environmental and Occupational Health 2007;62:5-11.
10. Malcoe LH, Lynch RA, Kegler MC, Skaggs VJ. Lead sources, behaviors, and socioeconomic factors in relation to blood lead of Native American and White children: a community-based assessment of a former mining area. Environmental Health Perspectives 2002;110 (Suppl 2):221-231.
11. Garry VF, Harkins M, Lyubimov A, Erickson L, Long L. Reproductive outcomes in the women of the Red River Valley of the north. I. The spouses of pesticide applicators: pregnancy loss, age at menarche, and exposures to pesticides. Journal of Toxicology and Environmental Health A 2002;65:769-786.
12. Boffetta P, Nyberg F. Contribution of environmental factors to cancer risk. British Medical Journal 2003;68:71-94.
13. Dubowsky SD, Suh H, Schwartz J, Coull BA, Gold DR. Diabetes, obesity, and hypertension may enhance associations between air pollution and markers of systemic inflammation. Environmental Health Perspectives 2006;114:992-998.
14. Fassett DW. Cadmium: biological effects and occurrence in the environment. Annual Review of Pharmacology 1975;15:425-435.
15. Goldberg MS, Burnett RT, Bailar JC, Tamblyn R, Ernst P, Flegel K, et al. Identification of persons with cardiorespiratory conditions who are at risk of dying from the acute effects of ambient air particles. Environmental Health Perspectives 2001;109(Suppl 4):487-494.
16. Mastin JP. Environmental cardiovascular disease. Cardiovascular Toxicology 2005;5:91-94.
17. Meliker JR, Wahl RL, Cameron LL, Nriagu JO. Arsenic in drinking water and cerebrovascular disease, diabetes mellitus, and kidney disease in Michigan: a standardized mortality ratio analysis. Environmental Health 2007;6:(serial online.)
18. Blair A, Sandler DP, Tarone R, et al. Mortality among participants in the agricultural health study. Annals of Epidemiology. 2005;15:279-285.
19. Jones CA, Parker TS, Ahearn M, Mishra AK, Variyam JN. Health status and health care access of farm and rural populations. US Department of Agriculture, Economic Information Bulletin Number 57, August 2009.

20. Hendryx M, O'Donnell K, Horn K. Lung cancer mortality is elevated in coal-mining areas of Appalachia. Lung Cancer 2008;62:1-7.

**Additional Information**

*See the Full Report that corresponds to this Brief for more detailed methods and findings from this study at: http://wvrhrc.hsc.wvu.edu/projects/2009/hendryx/*