A geographical information system-based analysis of cancer mortality and population exposure to coal mining activities in West Virginia, United States of America

Michael Hendryx[1], Evan Fedorko[2], Andrew Anesetti-Rothermel[3]

[1]Department of Community Medicine, P.O. Box 9190, West Virginia University, Morgantown, WV 26506, USA; [2]West Virginia GIS Technical Center, Department of Geology and Geography, West Virginia University, Morgantown, WV 26506, USA; [3]Mary Babb Randolph Cancer Center in partnership with the Prevention Research Center, Department of Community Medicine, P.O. Box 9190, West Virginia University, Morgantown, WV 26506, USA

Abstract. Cancer incidence and mortality rates are high in West Virginia compared to the rest of the United States of America. Previous research has suggested that exposure to activities of the coal mining industry may contribute to elevated cancer mortality, although exposure measures have been limited. This study tests alternative specifications of exposure to mining activity to determine whether a measure based on location of mines, processing plants, coal slurry impoundments and underground slurry injection sites relative to population levels is superior to a previously-reported measure of exposure based on tons mined at the county level, in the prediction of age-adjusted cancer mortality rates. To this end, we utilize two geographical information system (GIS) techniques – exploratory spatial data analysis and inverse distance mapping – to construct new statistical analyses. Total, respiratory and "other" age-adjusted cancer mortality rates in West Virginia were found to be more highly associated with the GIS-exposure measure than the tonnage measure, before and after statistical control for smoking rates. The superior performance of the GIS measure, based on where people in the state live relative to mining activity, suggests that activities of the industry contribute to cancer mortality. Further confirmation of observed phenomena is necessary with person-level studies, but the results add to the body of evidence that coal mining poses environmental risks to population health in West Virginia.

Keywords: mining, coal, cancer, mortality, geographical information system, Virginia.

Introduction

Cancer mortality in West Virginia and other parts of Appalachia is high relative to the United States of America as a whole (Huang et al., 2002; Halverson et al., 2004; Wingo et al., 2008). The higher cancer mortality in the region has been attributed to behavioural risks such as smoking, poor socio-economic conditions and problematic access to medical care (Huang et al., 2002). However, recent research evidence also points to the impact of the coal mining industry on population health. Persons who live in coal mining counties of Appalachia, compared to non-mining counties or the nation, have elevated all-cause (Hendryx, 2008; Hendryx and Ahern, 2009) and lung cancer (Hendryx et al., 2008) mortality, after controlling for socio-economic, health services and behavioural variables.

Coal contains many established carcinogens including arsenic, cadmium, chromium, nickel, beryllium and others (WVGES, 2007a), and coal extraction, processing and transportation activities have contaminated trillions of gallons of

Corresponding author:
Michael Hendryx
Department of Community Medicine
PO Box 9190, West Virginia University
Morgantown, WV 26506, USA
Tel. + 1 304 293 9206; Fax + 1 304 293 6685
E-mail: mhendryx@hsc.wvu.edu

water (Todd, 2008) and released tons of particulate matter into the air near mining communities (Ghose, 2007). Previous studies of the effects of coal mining on community health have relied on the tons of coal mined at the county level as the exposure measure, which is then correlated to health outcomes, also usually at the county level (Hendryx, 2008, 2009; Hendryx et al., 2008). Although statistical controls are made for confounding factors such as race/ethnicity, smoking, education, physician supply, poverty and others, the results are still limited to this exposure measure and have not accounted for more specific activities of the mining industry in relation to population concentrations and health outcomes.

The purpose of the current study was to use geographic information systems (GIS) to develop and test a more refined measure of population exposure to components of the mining industry in West Virginia compared to the exposure measure used in previous studies, by relating the comparative exposure measures to cancer mortality. The more refined measure is a distance-weighted population exposure score, distance being distance to components of the mining industry, as described in the next section. If it is true that exposure to extraction and processing activities of the mining industry increases cancer risk among community residents, one would expect distance-weighted population exposure measures to be more highly correlated to cancer mortality rates than the previous measure of tons of county-level coal mining. If, on the other hand, cancer mortality is not causally related to exposure to mining activity, but reflects only socio-economic or behavioural confounds, there should be no improvement in the capacity of the exposure measure to account for mortality rates. We have tested the hypotheses that (i) age-adjusted county cancer mortality rates will be positively associated with distance-weighted population exposure to coal extraction and processing activities, and (ii) the distance-weighted exposure measure will be more strongly correlated to cancer mortality than exposure based on tons of coal mined in the county.

Materials and methods

Study area

This study focuses on one state within the Appalachian region, West Virginia. Uniquely among Appalachian states, all 55 of West Virginia's counties fall within the Appalachian region as defined by the Appalachian Regional Commission (ARC, 2007). US Census data indicate that the total population of West Virginia is slightly over 1.8 million people, 37th among all states (US Census 2009a, b).

West Virginia is largely rural. Of the 55 counties, eight are classified as "metropolitan" by the US Department of Agriculture (USDA) Economic Research Service's Urban Influence Coding system (ARF, 2006). With the exception of Jefferson county, which lies in the eastern panhandle close to Washington D.C., the major urban area of West Virginia is the Kanawha River Valley, in the southwestern part of the state. The seven counties in this valley contain the state's two largest cities – Charleston, the state capital, and Huntington. The remaining 47 counties are all classified as "nonmetropolitan" and their respective codes cover the gamut from "micropolitan adjacent to large metro area" to "noncore adjacent to micro area, containing a town of 2,500–9,999 residents." In terms of population, the rural/urban dichotomy is reflected in the Census 2000 values. Approximately 50% of West Virginia's population resides within just 11 counties. Ten counties have less than 10,000 residents. West Virginia's counties have an average population density of 94.9 persons, and a median of 51.1 persons, per square land mile. When the county population density is mapped by natural breaks (JENKS) classification, 29 of West Virginia's 55 counties fall into the lowest value group.

West Virginia has a long history of coal mining as a principal industry. While most (if not all) West Virginia counties have been touched in some way by coal mining, most production has taken place within a few core areas. The coal infrastructure data utilized in this analysis is concerned with the location
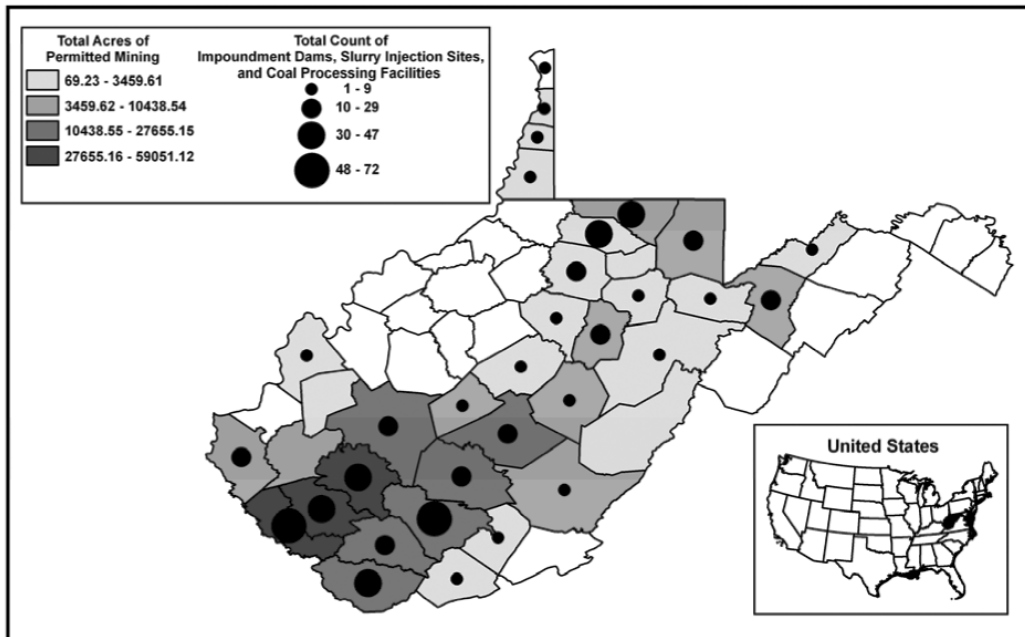
Fig. 1. West Virginia mining activities by count (1979-2004).

of various types of facilities (see the Data subsection below). As can be seen in Figure 1, these facilities are mostly located along a southeast (SW) to northeast (NE) trending band of counties through the central part of the state.

Data

Cancer mortality rates were taken from the Centers for Disease Control and Prevention (CDC) (Atlanta, GA, USA) public data website (CDC, 2008). The rates are age-adjusted using the 2000 US standard population, and are found for West Virginia counties as the rate per 100,000 person-years for 1979-2004. The person-year approach allowed us to aggregate across years to estimate cancer mortality in rural, less populated counties that typify most coal mining locations. Based on International Classification of Disease (ICD)-9 (years 1979-1998) and ICD-10 (years 1999-2004) diagnostic classifications, mortality rates were found for digestive, genital, urinary, breast, oral, respiratory, other and combined cancer sites. These

are the major diagnostic classifications provided on the CDC data consistently across ICD-9 and ICD-10 specifications and represent the major body systems for which rates are reported. Table 1 summarises the diagnostic categories.

US Census data for the year 2000 were used to find the population of each census block group within the state. This dataset was acquired from a DVD of base layer geographic information of the United States of America. The DVD was published by Environmental Systems Research Institute (ESRI) (Redlands, CA, USA), but the dataset itself originates from the United States Census Bureau.

Potential county-level covariates were based on prior studies (Hendryx and Ahern, 2008; Hendryx et al., 2008.) The time period, represented by covariates, was sometimes based on the 2000 Census, and sometimes on more recent estimates when available. These covariates include average poverty rate for 2000-2002, high school and college education rates in 2000, supply of primary care physicians per 1,000 population in 2001, smoking rate in 2003, percent race/ethnicity categories as of

Table 1. ICD-9 (years 1979-1998) and ICD-10 (years 1999-2004) codes used to classify cancer mortality by site.

ICD-9 Codes	ICD-9 Labels	ICD-10 Codes	ICD-10 Labels
140-149	Lip, oral cavity and pharynx	C00-C14	Lip, oral cavity and pharynx
150-159	Digestive organs and peritoneum	C15; C16; C18-C21; C22; C25	Esophagus; stomach; colon, rectum and anus; liver and intrahepatic bile duct; pancreas
160-165	Respiratory and intrathoracic	C32; C33-C34	Larynx; trachea, bronchus and lung
174-175	Breast	C50	Breast
179-187	Genital organs	C61; C53; C54-C55; C56	Prostate (males); cervix uteri; corpus uteri and uterus; ovary (females)
188-189	Urinary organs	C64-C65; C67	Kidney and renal pelvis; bladder
200-203; 204-208; 170-173; 190-199	Other malignant neoplasms of lymphatic and hematopoietic tissue; leukemia; all other and unspecified	C81; C82-C85; C88-C90; C96; C91-C95; C43; C70-C72	Hodgkin's disease; non-Hodgkin's lymphoma; multiple myelomas and immunoproliferative neoplasms; other unspecified neoplasm of lymphoid hematopoietic and related tissue; leukemia; all other and unspecified

2002 (i.e. White, Native American, African American, Asian American, non-white Hispanic), 2000 rural-urban continuum code, health insurance rate in 2000, and percent female population in 2000. Data for most covariates were obtained from the Area Resource File (ARF, 2006). Smoking rates were taken from the CDC Behavioral Risk Factor Surveillance System (BRFSS) website (CDC, 2007) supplemented with examination of the state public health department website.

Geographic data on activities of the coal mining industry included mining permit boundaries for mining sites, the point locations of surface slurry impoundment dams, the point locations of permitted underground injection sites, and the point locations of coal processing facilities (preparation plants). Most of the data compilation and manipulation was performed using ArcView GIS software, versions 9.2 and 9.3 (ESRI, Redlands, CA, USA). Most datasets originate from the West Virginia Department of Environmental Protection (WVDEP), but in some cases, additional analysis has been undertaken. The WVDEP publishes a spatial database of mine permit boundaries within West Virginia via an internet website (http://gis.wvdep.org/data/omr.html) (WVDEP,

2009). For the purpose of this study, we removed permit boundary polygons whose permit dates did not interest the time period of interest – 1979 to 2004. In total, we included 2,924 mining boundaries in the study. It is important to note that these boundary areas can overlap and thus, it is difficult to count a single boundary area as a single mine as certain mine features – such as roads – are often reused and re-permitted in subsequent mining operations. Likewise, the measure of total acres of mining per county is best described as approximate given this reality and the inherent inaccuracies of the data and it's reflection of reality on the ground.

Mining operations often involve the capture of used water in artificial impoundments, held in place by earthen dams, for the purpose of removing contaminants and non-combustibles. Acid mine drainage (AMD) is also held in surface impoundments. These coal impoundment dams are also regulated by the WVDEP. According to WVDEP GIS management, the overall spatial accuracy of the coal impoundment dam dataset utilized in this study is high (M Shank, WVDEP, personal communication). The data are difficult to maintain, however, and as such, temporal accuracy is difficult to quantify, but assumed to be less than current. In our own explo-

rations of the data, we compared the points to 2003 aerial photography and found that most (115 points, or 84%) of the points align with a dam in existence as of that date. A few (22 points, or 16%) appear to be reclaimed, although health considerations from these prior impoundments may still be relevant. At this time we cannot assign a reclamation date to these points. We used all 137 coal impoundment dams in this study.

We identified two sources of spatial information for coal processing facilities. The first, derived from the US Environmental Protection Agency (EPA)'s facility registry system, appeared to be incomplete when compared to the data obtained from the WVDEP. In the end, we derived a subset of points from the WVDEP's point database of coal activities permits that pertained to coal processing facilities. We then performed a basic photo alignment of these features using 2003 aerial photography. Where possible, the point was relocated to a more accurate location. Of 76 entities, 46 were realigned in this way. The remaining 30, most of which had been reclaimed at the time of the aerial photography, remained in their originally published location.

The last coal mining feature utilized in this study is slurry injection site locations. These are areas where waste water from mining, drilling or processing has been injected into underground voids for the purpose of storage. These entities are permitted by the WVDEP as part of the mine process and the points at which they discharge (into a nearby stream or reservoir) are permitted by the EPA as National Pollution Discharge Elimination System (NPDES) points. The content of the slurry is monitored at the discharge points as part of the NPDES regulatory process. The data we use in this study were compiled from the WVDEP permit database and EPA NPDES database. This study did not differentiate between injection sites and discharge points. In total, the dataset contains 270 points.

The analysis was completed in three phases. First, we conducted an exploratory spatial data analysis (ESDA) to determine if, in fact, any quantifiable spatial relationships existed between our existing data. Second, we developed a distance-based index describing, per county, the proximity of that county's population to coal mine features. Last, we utilized this index in a regression analysis, discussed in more detail in the "Analysis and Results" section.

Analysis

The current ESDA of the association between coal extraction and processing activities and cancer mortality rates was conducted on census block group and county level data for West Virginia. GIS and its spatial analysis tools were used to examine the spatial association/autocorrelation between county-level age-adjusted combined cancer mortality rates and total tonnage of coal production per county. First, a global measure of spatial autocorrelation to measure the level and direction (e.g. positive or negative) of association for the entire sample was used. Here, we tested the hypothesis that the univariate and bivariate global measures result in positive spatial autocorrelation. If the variables of interest are found to have positive spatial autocorrelation, then their local statistics of spatial autocorrelation are calculated to identify where these spatial clusters exist.

Global spatial autocorrelation was used to test the overall spatial dependency of the variables of interest by calculating a Moran's *I* statistic. Moran's *I* is defined as:

$$I = \frac{\frac{N}{\sum_i \sum_j \sum_{ij}} \sum_i \sum_j W_{ij} Z_i Z_j}{\sum_i Z_i^2}$$

where Z_i is the deviation of the variable of interest with respect to the mean; W_{ij} the spatial weight matrix (i.e. a binary matrix with "ones" in position *i,j* whenever observation *i* is a neighbour of observation *j*, and otherwise "zero"); N the number of observations and $\sum_i \sum_j W_{ij}$ the standard deviation (Anselin, 1995). The resulting test is similar to a correlation coefficient as it varies between -1.0 and +1.0. This test of global spatial autocorrelation was computed using GeoDa 0.9.5-i software down-

loaded from Arizona State University's GeoDa Center.

The variables of interest, county-level age-adjusted combined cancer mortality rates (1979-2004) and tonnage of coal production per county (1986-2005), were joined to a georeferenced spatial county layer file of West Virginia and its first order (neighbouring) counties from surrounding states. The tonnage measure, rather than the distance-weighted measure, was used for the spatial analysis because we could capture tonnage measures from border counties outside West Virginia. However, the mapping of mining activities was available only within the state. The data were joined to the georeferenced spatial county layer file by their unique 5-digit federal information processing standards (FIPS) code. A spatial weight (W) was then created to impose a neighbourhood structure on the data. We chose and created a queen contiguity weight matrix for each variable of interest.

We then calculated two univariate Global Moran's I statistics, one for the county-level age-adjusted combined cancer mortality rates (1979-2004) and one for the total tonnage of coal production per county (1986-2005). We also calculated a bivariate Global Moran's I statistic with the total tonnage of coal production per county (1986-2005) along the X-axis and the county-level age-adjusted combined cancer mortality rates (1979-2004) along the Y-axis. Lastly, we randomized each Global Moran's I statistic by 999 random permutations to yield a pseudo p-value. Finally, we interpreted Moran's I by whether the values of I were significantly greater than the expected values (positive spatial autocorrelation) or significantly less than the expected values (negative spatial autocorrelation).

The local indicators of spatial autocorrelation (LISA) statistic has been described, in the words of Anselin (1998), as "an indicator that achieves two objectives: it allows for the detection of significant patterns of local spatial association (i.e. association around individual location), and it can be used as a test for stability of a global diagnostic (i.e. to assess the extent to which the global pattern of association is reflected uniformly throughout the data set".

This local Moran's I statistic is derived form the Global Moran's I statistic by the formula:

$$I = \frac{Z_i}{m_2} \sum_j W_{ij} Z_j, \text{ where}$$

$$m_2 = \sum_i Z_i^2, \text{ hence}$$

$$I = \sum_i \frac{I_i}{N},$$

where N is the number of observations (Anselin, 1995). The current LISA analysis is identical to the global spatial autocorrelation analysis in software, data, spatial weight matrix, number of tests and permutations. However, this analysis results in the production of a 95% level of significance cluster map, which classifies four different cluster types: high-high, low-low, low-high and high-low (Anselin, 2003). The high-high and low-low clusters suggest areas where similar values are clustered together indicating positive spatial autocorrelation, while low-high and high-low clusters indicate spatial outliers and negative spatial autocorrelation (Anselin, 2003; GeoDa Center, 2009).

Distance-weighted at-risk population index

The results of the ESDA indicate that a quantifiable relationship between population proximity to coal mine features and cancers exists (see the "Analysis and Results" section). We were presented with a challenge at this point: although we are fortunate to have access to point data that precisely locate coal mine features, our cancer rate information remained at a coarse county-level resolution. We opted to create a simple distance-based statistic using the finest scale data available which could then be summed at the county level. While the potential exposure variable would still be coarse, the finer scale of the source data should enhance the analysis and allow for a test of the primary study

hypothesis. We opted to utilize distance to account for the potential fallacies that exist when quantifying exposure based on counting facilities based on arbitrary boundaries (Maantay, 2002).

We developed the refined exposure index using simple GIS techniques, the main steps of which are illustrated in Figure 2. As described above, the polygon location of every coal mine was mapped, along with the point location of every coal preparation plant, slurry impoundment dam, and permitted underground slurry injection site. Next, we determined Euclidean distance from each of the coal mining infrastructure entities via 30x30 m grid cell distance measures. Within each of the study area's census block groups, we calculated the mean distance in km from each to the nearest (i) mine boundary poly-

gon, (ii) impoundment dam centroid, (iii) injection site, and (iv) preparation plant (Fig. 2, step 1). We then found the inverse distance (1/distance) for each mine infrastructure type for each block group. The mean inverse distances were multiplied by the population of the block group. This results in a value per block group/infrastructure type wherein closer distances and bigger populations have larger values, and farther distances and smaller populations have smaller values (step 2).

These values are representative of a population "P" variable adjusted by inverse distance as a metric of exposure. These values were summed across type "t" (mine "m", impoundment dam "d", injection site "I", and preparation plant "p") (step 3) and across block group "k" to the county-level (step
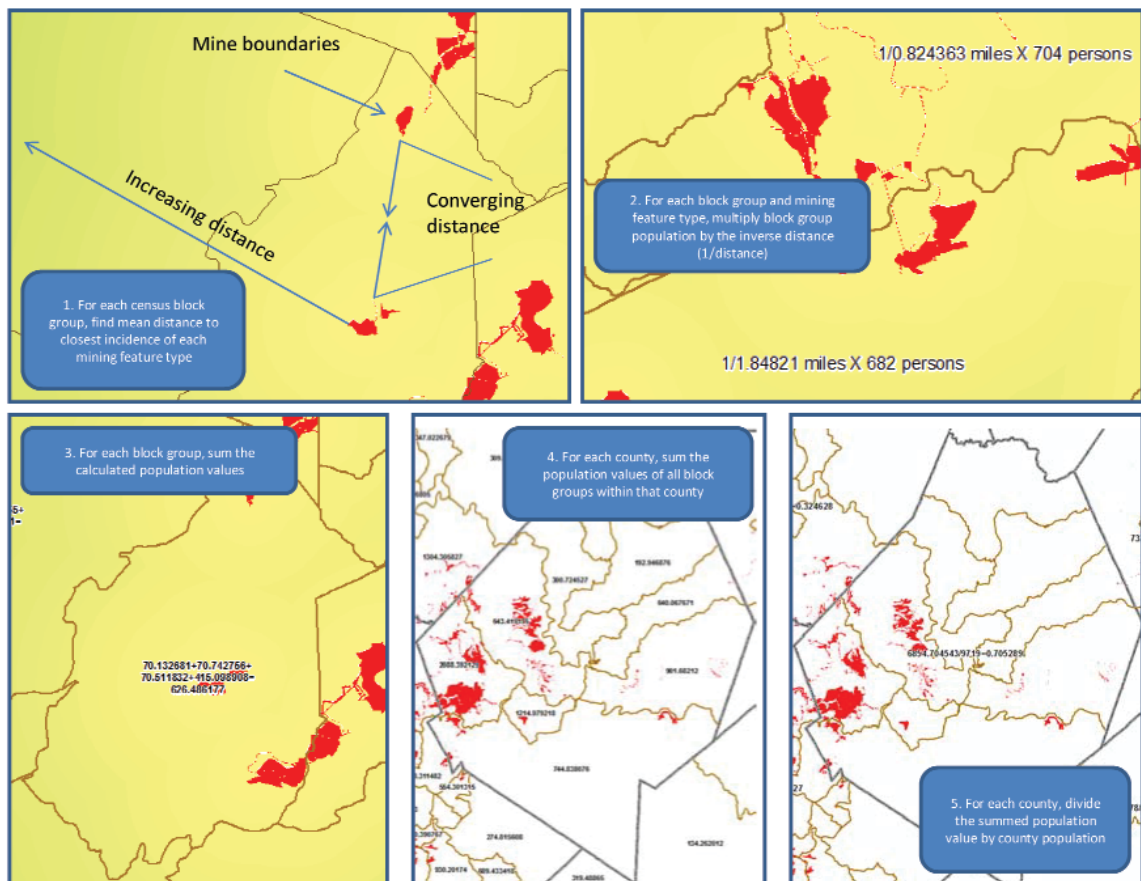
Fig. 2. Representation of construction of the distance-weighted at risk population index.

4), and divided by the total county population "Pc", to find the total mean per capita county-level distance-weighted mining impact score (step 5). This measure of exposure can be expressed as:

$$\frac{\sum_{i=1}^{k}(\sum P_k\,(l/d_t))}{P_c}\text{ , where}$$

$$\sum P_k\,(l/d_t) = P_k\,/\,d_m + P_k\,/\,d_d + P_k\,/\,d_i + P_k\,/\,d_p$$

As with previously published studies (Hendryx et al., 2007, 2008ab, 2009; Hendryx and Ahern, 2008) the "old" measure of mining exposure was constructed by creating a 3-level variable that divided counties into no mining (n = 24 counties), moderate mining levels (n = 16), or high mining levels (n = 15), based on tons of coal mined at the county-level. The boundary between moderate and high was set at 4 million tons for the years 1986-2005, the median value dividing mining counties into equal sized groups. County-level coal production statistics were drawn from the West Virginia Geologic and Economic Survey (WVGES, 2007b) which includes statistics only back to 1986. The new and old measures of exposure are termed the distance-weighted exposure measure and the tonnage exposure measure, respectively.

Results

Analysis

All three global spatial autocorrelation tests (cancer, coal and the bivariate) resulted in positive spatial autocorrelation. The univariate global spatial autocorrelation test of county-level age-adjusted combined cancer mortality rates yielded: Moran's $I = 0.405$; $E[I] = -0.011$; mean = -0.008; SD = 0.066, where: $E[I]$ is the expected Moran's I and SD is the standard deviation of Moran's I. The univariate global spatial autocorrelation test of county-level total tonnage of coal production yielded: Moran's $I = 0.354$; $E[I] = -0.011$; mean = -0.011; SD = 0.061.

The bivariate global spatial autocorrelation between total tonnage of coal production per county and county-level age-adjusted combined cancer mortality rates yielded: Moran's $I = 0.218$; $E[I] = -0.011$; mean = -0.001; SD = 0.049. All three global spatial autocorrelation tests yielded a pseudo significance value of P <0.001, which indicates that the data are not spatially random.

The LISA tests resulted in three 95% significance cluster maps (Fig. 3). The county-level, age-adjusted, combined cancer mortality rates' 95% significant cluster map (Fig. 3a) yielded six counties (Wayne, Lincoln, Mingo, Logan, Boone and Wyoming) categorized as being high-high clusters and seven counties (Preston, Tucker, Randolph, Pocahontas, Pendleton, Grant and Hardy) classified as being low-low clusters. Two counties, Putnam (low-high) and Mineral (high-low), were found to be spatial outliers. Spatial clusters, high-high and low-low, are said to be significantly higher and lower than the global Moran's I, while the low-high and high-low clusters are individual locations significantly different than their neighbors.

The county-level, total tonnage of coal production's 95% significant cluster map (Fig. 3b) yielded six counties (Marshall, Mingo, Logan, Boone, Wyoming and Raleigh) as being high-high clusters and nine counties (Jefferson, Berkeley Hardy, Pendleton, Monroe, Pleasants, Ritchie, Wirt, and Gilmer) classified as low-low clusters. Two counties, Wetzel and Lincoln were found to be low-high spatial outliers. The bivariate LISA 95% significant cluster map between county-level, total tonnage of coal production and county-level, age-adjusted combined cancer mortality rates (Fig. 3c), yielded five counties (Mingo, Logan, Boone, Wyoming and Raleigh) that were classified as high-high clusters, while eight counties (Wetzel, Pleasants, Ritchie, Wirt, Gilmer, Hardy, Pendleton and Monroe) were classified as being low-low clusters. There were three low-high classified spatial outliers: Jefferson, Berkeley, and Lincoln, while Marshall was the only high-low classified spatial outlier.
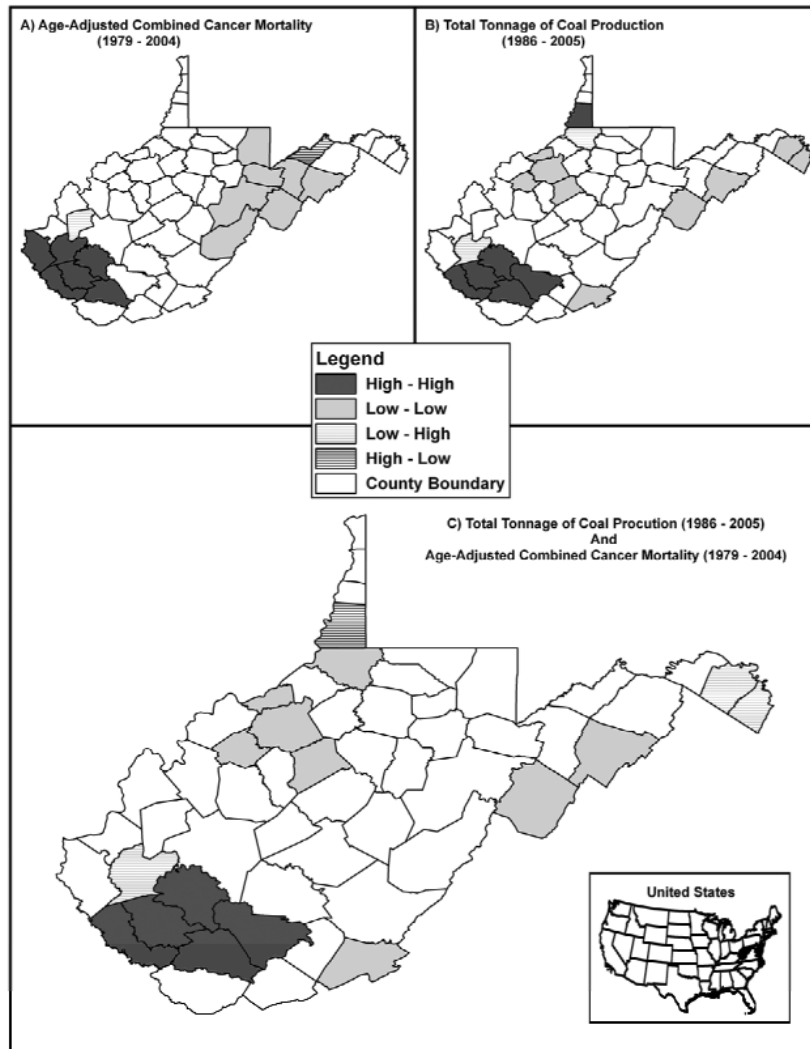
Fig. 3. Local indicators of spatial autocorrelation (LISA) cluster maps.

Correlation and regression analysis

Correlations were found between age-adjusted cancer mortality rate and the alternative exposure measures in comparative models. We used Pearson correlations for the distance-weighted exposure measure and Spearman correlations for the categorical tonnage measure. Then, ordinary least squares regression models were used to compare the distance-weighted and tonnage exposure measures while accounting for covariates within constraints of the limited sample size, with age-adjusted cancer mortality as dependent variables. Covariates were identified using preliminary stepwise regression models setting entry into the model at P <0.05.

Descriptive statistics on mining activity and cancer mortality are shown in Table 2. The average person lived 7 km from the nearest coal mine, 27 km from the nearest preparation plant, and 32 km from the nearest slurry impoundment or permitted underground injection site.

Table 2. Descriptive statistics: age-adjusted cancer mortality rates per 100,000 person-years and mining activity.

Cancer statistics	Mean*	Standard deviation (SD)*
Breast	14.8	2.5
Digestive	48.1	4.9
Genital	22.9	2.4
Oral	2.7	1.0
Respiratory	66.0	11.6
Urinary	9.0	1.7
Other	30.3	4.5
Total	193.6	18.5

*County mortality rates

Mining statistics	N	Mean km distance within census block (SD)	Minimum-maximum
Injection sites	270	32.6 (31.4)	0.6 – 131.9
Slurry impoundments	105	32.1 (33.0)	0.9 – 143.4
Mining sites	4,026	7.0 (6.8)	0.08 – 35.1
Preparation plants	76	27.0 (22.6)	1.0 – 114.3

Table 3 shows that the correlations between cancer mortality and the total distance-weighted exposure were higher than the corresponding correlations between cancer mortality and the tonnage exposure measure for all cancer sites except breast cancer. Both the old and the new measures of exposure were correlated significantly to "other" and total cancer, but the correlations reached higher levels of statistical significance for the new measure. Respiratory cancer was found to be correlated to the distance-weighted measure but not to the tonnage measure.

Selection of covariates for regression analysis is limited by the small sample size. Stepwise regression models were used to identify the most important independent variables based on a P <0.05 inclusion criterion. Two variables were thus identified: smoking rates and the weighted exposure variable. This analysis used total cancer mortality rates as the dependent variable. No other variable, including the tonnage exposure measure, race/ethnicity percentages, education, poverty, primary health care access, health insurance rates, rural-urban setting or percent female population contributed significant additional variance beyond exposure and smoking when considered in stepwise fashion. However, to test the second hypothesis, the tonnage exposure measure was carried forward with the distance-weighted exposure measure and the smoking variable into the next analyses.

When the two exposure variables were compared controlling for smoking, total cancer mortality rate remained independently associated with the dis-

Table 3. Correlations between mining activity and cancer mortality rates for distance-weighted exposure (Pearson correlation) and tonnage exposure (Spearman correlation) measures.

Cancer site	Components of distance weighted measure				Distance weighted measure	Tonnage measure
	Injection sites	Impoundments	Mines	Prep plants	Total exposure	Categorical coal mining in tons
Breast	0.28*	0.33**	0.26	0.30*	0.29*	0.29*
Digestive	0.01	0.08	0.13	-0.01	0.11	-0.06
Genital	-0.2	-0.09	0.01	-0.07	-0.03	-0.13
Oral	-0.03	-0.01	0.07	0.09	0.06	0.03
Respiratory	0.23	0.37***	0.57****	0.36***	0.53****	0.24
Urinary	0.01	0.07	0.08	0.01	0.07	-0.04
Other	0.34**	0.34**	0.43****	0.41***	0.44****	0.38***
Total	0.24	0.37***	0.55****	0.36***	0.51****	0.28*

* P <0.05; ** P <0.02; *** P <0.01; **** P <0.001

Table 4. Regression results to predict age-adjusted cancer mortality from alternative exposure specifications, controlling for smoking rates.

	Distance weighted measure			Tonnage measure		
	Standardized Beta	P <	Model adjusted R^2*	Standardized Beta	P <	Model adjusted R^2*
Cancer site						
Breast	0.303	0.04	0.05	0.283	0.04	0.05
Respiratory	0.367	0.002	0.42	0.228	0.05	0.35
Other	0.441	0.002	0.16	0.324	0.02	0.1
Total	0.369	0.003	0.36	0.206	0.09	0.28

* Model adjusted R^2 includes exposure measure and smoking rate as independent variables. F values for all models significant at P <0.02 or greater, with the exception of both models for breast cancer, which had P <0.10.

tance-weighted exposure measure but not the tonnage measure. Effects were significant (P <0.05) for both exposure measures for breast, respiratory, and "other" cancer, but reached higher significance levels for respiratory and "other" cancer. These results are summarized in Table 4. Results for the remaining cancer sites were not significant with either exposure measure and are not included in the table.

Discussion

Results of the correlation and regression analyses supported the study hypotheses. The distance-weighted, at-risk population coal mining exposure measure was significantly correlated to cancer mortality in West Virginia. For total cancer and three cancer subgroups, the exposure measure was correlated to higher mortality after controlling for smoking rates. The previous exposure measure, based on tonnage, was not related as strongly to cancer mortality. For total cancer the effect was significant only for the distance-weighted measure. The data are correlational and causal links cannot be proven, but the superior performance of the distance-weighted exposure measure is consistent with the possibility of environmental contamination from the mining industry as a causal factor in the etiology of cancer for populations residing in West Virginia.

The global Moran's *I* indicated that the variables of interest had positive spatial autocorrelation and were spatially dependent. However, as with all correlation studies, other confounding factors could have been spatially correlated to one or both variables of interest. The LISA statistic identified these positive spatial clusters (high-high and low-low), while the spatial outliers identified areas for future research and/or investigations.

In previous published studies, based on the county measure of tonnage as the exposure variable, significant effects were found between this measure and various health outcomes including lung cancer (Hendryx et al., 2008), chronic heart, lung and kidney disease (Hendryx, 2009), and self-reported health (Hendryx and Ahern, 2008). However, these previous studies considered larger samples of counties representing the entire Appalachian region or the whole nation. In one case it was based on a large person-level sample of over 16,000 cases. In the current study, that was limited to a small sample of 55 counties, the smoking-adjusted associations did not reach as strong a statistical significance with the old categorical variable as with the new distance-weighted exposure variable as the new measure has stronger power to detect effects.

All four components of the mining industry (injection sites, preparation plants, impoundment ponds and mines) were related to one or more cancer types, although the injection sites were the least correlated and mining boundaries the most strongly correlated to cancer. This pattern is particularly

clear for respiratory and total cancer, and as respiratory cancer is the most common fatal cancer (see Table 1), it drives much of the total cancer variance. The strong association between respiratory cancer and mining boundaries, controlling for smoking, may reflect air quality problems around the mines, especially at mountaintops and other surface mining operations. Dust from coal mining is more severe at surface versus underground coal mining sites (Ghose, 2007) and surface mining as a percent of total mining has been increasing in West Virginia (WVGES, 2007b).

We recognize that the study has several limitations. Firstly, imprecision exists in the temporal relationships between exposure and cancer mortality. The development of cancer from exposure to environmental pollutants is a long-term phenomenon, and yet in this study mortality for the years 1979-2004 was related to activities of the industry as they could be constructed from available sources for approximately the same years. Cancer mortality was collapsed across these years to produce sufficient cases to examine specific types of cancer at the county-level in a relatively rural state with a small population. We assume that current mining reflects past mining, and this assumption is justified based on the long history of coal mining in the region (Hickam, 1998; Goodell, 2006) but, to the extent this not being the case, error in the exposure estimates is introduced. However, because the study compared alternative measures of exposure, each with this time limitation, the relatively stronger results for the distance-exposure measure are still relevant.

The temporal imprecision extends to estimates of population characteristics that are taken from the 2000 Census and nearby years, relative to the long aggregation of cancer mortality over the period 1979-2004. Changes in population characteristics due to migration or other dynamics might influence cancer estimates separately from mining activities. We note two population characteristics in response to this. One is that the overall population of West Virginia declined between the 1980 and 1990

Census (1.94 million to 1.79 million), then was largely stable from 1990 to 2000 (1.79 to 1.81 million.) The second is that population loss to emigration affected coal-mining counties significantly more than non-mining counties: between 1980 and 1990, the average coal mining county lost 5,233 people to migration compared to a loss of 1,175 people for non-mining counties (WVDHHR 2002). Between 1990 and 2000, the average coal mining county lost 663 people, while the average non-mining county gained 2,061 people (WVDHHR, 2002). The differential loss from mining areas could serve to make our observed mining effects more conservative than they are, because people may become exposed in mining area but develop cancer later in another area.

Secondly, although these methodologies are in line with other studies (Lin et al., 2002; Reynolds et al., 2005; Choi et al., 2006) we recognize the limitations of simple distance proximity equations (Maantay, 2002). More detailed methods, such as the development of pollution surfaces (Hoek et al., 2001; Buzzelli and Jerrett, 2003), the creation of dispersion models (English et al., 1999; Poulstrup and Hansen, 2004), or more advanced proximity analysis may yield better results. We utilized the best data available to complete this study. Many of the mentioned methods would require much more detailed data, such as particulate matter measurements or water sampling regimes, or referring to data at the person level rather than the county level such as individual cancer data. The results of this study, though limited by methodology, should be sufficient to justify more detailed field studies.

Thirdly, results of the correlation and regression analyses assumed independence between observations, when in fact the results of the spatial models demonstrated autocorrelation between adjacent counties in the dependent variable, cancer mortality rates. Models that account for spatial autocorrelation such as linear mixed models might be preferable but are constrained by the data (i.e. a sample of 55 at the single level of the county). These correlated observations may be expected to overestimate

effects. However, as noted above, this problem would affect both the tonnage measure and the distance-weighted measure, and so the relatively better performance of the distance-weighted measure is still pertinent.

Lastly, the study is also limited by the ecological design in that no direct, person-level measures of environmental exposure are available. The small sample size, 55 counties, provides limited statistical power to detect effects, and the lack of effects for some cancer types may be a consequence of this. Follow up research is needed to verify environmental impacts of coal mining on local air and water quality, and relate these impacts to population health by examining person-health exposures and health outcomes.

References

Anselin L, 1995. Local Indicators of Spatial Association-LISA. Geogr Anal 27, 93-115.

Anselin L, 1998. GIS research infrastructure for spatial analysis of real estate markets. J Hous Res 9, 113-133.

Anselin L, 2003. GeoDa 0.9 User's Guide. Urbana, IL, Department of Agricultural and Consumer Economics, University of Illinois, USA, pp. 99-105.

ARC, 2007. Appalachian Regional Commission. Retrieved 08-17-07, from http://www.arc.gov/index.jsp.

ARF, 2006. Area Resource File. Rockville, MD, U.S. Department of Health and Human Services Health Resources and Services Administration, Bureau of Health Professions.

Buzzelli M, Jerrett M, 2003. Comparing proximity measures of exposure to geostatistical estimates in environmental justice research. Environ Hazards 5, 13-21.

CDC, 2007. Behavioral Risk Factor Surveillance System. Centers for Disease Control and Prevention. Retrieved 07-11-07, from http://www.cdc.gov/brfss/index.htm.

CDC, 2008. Compressed Mortality File. Centers for Disease Control and Prevention. Retrieved 04-30-08, from http://wonder.cdc.gov/mortSQL.html.

Choi HS, Shim YK, Kaye WE, Ryan PB, 2006. Potential residential exposure to Toxics Release Inventory chemicals during pregnancy and childhood brain cancer. Environ

Health Perspect 114, 1113-1118.

English P, Neutra R, Scalf R, Sullivan M, Waller L, Zhu L, 1999. Examining associations between childhood asthma and traffic flow using a Geographic Information System. Environ Health Perspect 107, 761-767.

GeoDa Center, San Diego State University, 2009. Pseudo P-Values. Glossary of Key Terms. Retrieved 03-02-2009, from http://geodacenter.asu.edu/node/390.

Ghose MK, 2007. Generation and quantification of hazardous dusts from coal mining in the Indian context. Environ Monit Assess 130, 35-45.

Goodell J, 2006. Big Coal. Boston, Houghton Mifflin, USA.

Halverson JA, Ma L, Harner EJ, 2004. An Analysis of Disparities in Health Status and Access to Health Care in the Appalachian Region. Washington, DC, Appalachian Regional Commission.

Hendryx M, 2008. Mortality rates in Appalachian coal mining counties: 24 years behind the nation. Environ Jus 1, 5-11.

Hendryx M, 2009. Mortality from heart, respiratory, and kidney disease in coal mining areas of Appalachia. Int Arch Occup Environ Health 82, 243-249.

Hendryx M, Ahern M, Nurkiewicz T, 2007. Hospitalization patterns associated with Appalachian coal mining. J Toxicol Environ Health 70, 2064-2070.

Hendryx M, Ahern M, 2008. Relations between health indicators and residential proximity to coal mining in West Virginia. Am J Pub Health 98, 669-671.

Hendryx M, Ahern M, 2009. Mortality in Appalachian coal mining regions: the value of statistical life lost. Public Health Rep 124, 541-550.

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

Hickam H 1998. October Sky. Dell Publishing, New York, USA.

Hoek G, Fischer P, van den Brandt P, Goldbohm S, Brunekreef B, 2001. Estimation of long term average exposure to outdoor air pollution for a cohort study on mortality. J Expo Anal Environ Epidemiol 11, 459-469.

Huang B, Wyatt SW, Bottorff D, Lingerich E, Hall HI, 2002. Cancer death rates - Appalachia, 1994-1998. MMWR Morb Mortal Wkly Rep 51, 527-529.

Lin G, Allen DE, Penning MJ, 2002. Examining distance effects on hospitalizations using GIS: a study of three health

regions in British Columbia, Canada. Environ Plan 34, 2037-2053.

Maantay J 2002. Mapping environmental injustices: pitfalls and potential of geographic information systems in assessing environmental health and equity. Environmental Health Perspectives 110(suppl 2), 161-171.

Poulstrup A, Hansen HL, 2004. Use of GIS and exposure modeling as tools in the study of cancer incidence in a population exposed to airborne dioxin. Environ Health Perspect 112, 1032-1036.

Reynolds P, von Behren J, Gunier RB, Goldber DE, Harnley M, Hertz A, 2005. Agricultural pesticide use and childhood cancer in California. Epidemiology 16, 93-100.

Todd J, 2008. A New Shared Economy for Appalachia: An Economy Built upon Environmental Restoration, Carbon Sequestration, Renewable Energy and Ecological Design, University of Vermont, USA.

US Census, 2009a. Population estimates (geographies ranked by estimates), Table GCT-T1-R. American Fact Finder. Retrieved 06-18-09, from http://factfinder.census.gov/.

US Census, 2009b. West Virginia - Fact Sheet - Census 2000.

American Factfinder. Retrieved 06-18-09, from http:// factfinder.census.gov/.

Wingo PA, Tucker TC, Jamison PM, Martin H, McLaughlin C, Bayakly R, Bolick-Aldrich S, Colsher P, Indian R, Knight K, Neloms S, Wilson R, Richards TB, 2008. Cancer in Appalachia, 2001-2003. Cancer 112, 181-192.

WVDEP, 2009. West Virginia Department of Environmental Protection. Retrieved 05-26-09, from http://www. wvdep.org/#.

WVDHHR, 2002. A look at West Virginia's population by decade, 1950-2000. Health Statistics Center Statistical Brief No. 8. West Virginia Department of Health and Human Services. Retrieved 04-07-10, from http://www.wvdhhr.org/ bph/oehp/hsc/briefs/eight/default.htm.

WVGES, 2007a. Trace elements in West Virginia Coals. West Virginia Geologic and Economic Survey. Retrieved 10-06-07, from http://www.wvgs.wvnet.edu/www/datastat/te/ index.htm.

WVGES, 2007b."The West Virginia Geological and Economic Survey. Retrieved 10-06-07, from http://www. wvgs.wvnet.edu/.