Detection of clusters of lung cancer incidence in Idaho with ecological adjustment for smoking, radon, fine particulate air pollution, elevation, and socioeconomic position.

July 2020

A Publication of the

ACKNOWLEDGMENTS

This report has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN261201800006I and the Centers for Disease Control and Prevention, Department of Health and Human Services, under Cooperative Agreement 1NU58DP006270. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute, Centers for Disease Control and Prevention, the Idaho Department of Environmental Quality, or the Idaho Department of Health and Welfare. The authors thank Racheal Castañeda and James Faust (Idaho Department of Health and Welfare) for providing data on residential radon test results and Christopher Murphy (Idaho Department of Health and Welfare) for providing Behavioral Risk Factor Surveillance System data on smoking prevalence.

SUGGESTED CITATION:
Johnson CJ, Willming M, Paden N. *Detection of clusters of lung cancer incidence in Idaho with ecological adjustment for smoking, radon, fine particulate air pollution, elevation, and socioeconomic position*. Boise, ID: Cancer Data Registry of Idaho; July 2020.

CANCER DATA REGISTRY OF IDAHO
P.O. Box 1278
Boise, Idaho 83701-1278
208-489-1380 (phone)
208-344-0180 (FAX)
http://www.idcancer.org

Abstract

Background

The purpose of this study was to better understand spatial patterns in lung cancer incidence in the U.S. state of Idaho using a combination of multilevel modeling of area-level risk factors and a cluster detection method.

Methods

Poisson multilevel mixed models were fitted to model the incidence counts of lung cancer at the geographical level of census tract. Six separate models of increasing complexity were estimated. For each set of model results, discrete Poisson scan statistics were calculated using SaTScan to identify spatial clusters.

Results

Accounting for population sizes by age group and sex, but not for additional risk factors, SaTScan identified thirteen statistically significant clusters. Using the estimated tract-level counts from the final model, SaTScan identified two statistically significant clusters of census tracts.

Conclusions

Most of the differences in census tract-level lung cancer incidence in Idaho were explained by area-level measurements of known risk factors.

Keywords

Cancer cluster; lung cancer; spatial analysis; air pollution; radon; tobacco use

Abbreviations

AGGIE	Automated Geospatial Geocoding Interface Environment
APC	annual percent change
BRFSS	Behavioral Risk Factor Surveillance System
CAWG	Idaho Cancer/Cluster Analysis Working Group
CDRI	Cancer Data Registry of Idaho
EPA	U.S. Environmental Protection Agency
LDS	Church of Latter-day Saints
NAD 1983	North American Datum 1983, the current geodetic datum used for North America
pCi/L	picocuries per liter
PM 2.5 $\mu g/m^3$	air pollutants with an aerodynamic diameter less than 2.5 micrometers, measured in micrograms per cubic meter of air
RR	relative risk
SAS GLIMMIX	statistical procedure used to fit generalized linear mixed models
SaTScan	software for the spatial, temporal, and space-time scan statistics
SEP	socioeconomic position
U.S.	United States of America

Background

In 2017, citizens requested analyses of three suspected lung cancer incidence geographic clusters in the U.S. state of Idaho, one each in the Boise, Lewiston, and Twin Falls metropolitan areas. Idaho's interagency Cancer/Cluster Analysis Working Group (CAWG) was chartered in 2003 in part to review summarized cancer data to identify clusters and to review and respond to cancer cluster concerns. CAWG utilizes guidelines for investigating non-infectious disease clusters [1] based on recommendations from the U.S. Centers for Disease Control and Prevention and the Council of State and Territorial Epidemiologists [2]. To address these concerns, we conducted an initial statewide analysis of lung cancer geographic patterns and communicated the findings to the concerned citizens. The present study builds upon that analysis by using lung cancer incidence data for the period 2007-2016 and taking into account ecological risk factors for lung cancer including tobacco smoking, residential radon exposure, fine particulate outdoor air pollution, elevation, and socioeconomic position (SEP).

Lung cancer is the leading type of cancer in the world in terms of incidence and mortality [3]. Lung cancer is the third leading type of cancer in terms of incidence in Idaho [4] and, as for all U.S. states, the leading type of cancer mortality [5]. In the U.S., lung cancer incidence rates generally increase with age and cases are uncommon in persons less than 50 years of age. Lung cancer incidence rates are higher in males than females, but the gap is narrowing due to increased smoking rates among women in recent decades [6]. In the state of Idaho, the 2014 age-adjusted incidence rate of lung cancer was 48.8 per 100,000 persons, 56.0 per 100,000 males and 43.0 per 100,000 females [7]. The national rates were 68.1 per 100,000 males and 50.8 per 100,000 females.

Major risk factors for lung cancer are well-established. Smoking tobacco products is the leading cause of lung cancer [8]. Radon is an odorless, colorless radioactive gas formed from the natural decay of uranium that is found in nearly all soils [9]. According to the U.S. Environmental Protection Agency (EPA), radon is the leading cause of lung cancer among non-smokers and overall the second leading cause of lung cancer [10]. The EPA estimates that radon is responsible for about 21,000 lung cancer deaths in the U.S. each year, including about 2,900 among people who have never smoked. People can be exposed to radon primarily from breathing radon in air that comes through cracks and gaps in the foundations of buildings and homes. The majority of the land mass of Idaho is designated by the U.S. Environmental Protection Agency as Radon Zone 1, where a higher proportion of residential radon concentrations are greater than 4.0 picocuries per liter and thereby confer greater risks [11]. Studies have provided evidence that air pollution from urban and industrial sources is an independent risk factor for lung cancer [12,13]. In the developed world, it is estimated that 70-90% of lung cancer deaths are attributable to tobacco smoking [14], 3-20% to residential radon exposure [15], and 5-13% to particulate air pollution [16]. In addition, these risk factors may interact with each other to yield higher cumulative risk. For example, a smoker who is also exposed to radon has a much higher risk of lung cancer [17]. Some studies have found differences in lung cancer incidence rate with elevation, postulating different reasons [18,19,20]. Krieger and colleagues have recommended that the percent of persons below poverty at the census tract level be used to measure the impact of socioeconomic position on health [21]. A prior study using incidence data from a large number of U.S. registries and poverty estimates from the U.S. Census Bureau's American Community Survey showed a monotonic relationship between census tract poverty category and lung cancer incidence rates [22].

To help better understand how these risk factors (smoking prevalence, residential radon, fine particulate outdoor air pollution, elevation, SEP) may be contributing to the rates of lung cancer in Idaho and the existence of cancer clusters, we assembled ecological data on the risk factors for all areas of the state and conducted an analysis of lung cancer incidence adjusting for differences in the risk factors. The purpose of this study was to

better understand spatial patterns in lung cancer incidence using a combination of multilevel modeling of area-level risk factors and cluster detection methods. Our study is similar to other studies that have used a combination of statistical models, cluster detection, and geographic information systems using ecological data to perform in-depth analysis or follow-up to cancer incidence or mortality cluster concerns [23,24,25]. This study builds on previous research of lung cancer geographic patterns by utilizing a more comprehensive set of ecological risk factors, including the interaction between residential radon concentration and smoking prevalence.

Methods

Geocoded lung cancer incidence data

We conducted the statistical analysis of geographic patterns in lung cancer incidence at the level of census tract, using U.S. Census 2010 tract boundaries. Census tracts are small, relatively permanent subdivisions of counties that contain, on average, about 4,000 people. Census tracts are designed to be relatively homogeneous units with respect to population characteristics, economic status, and living conditions. Counts of incident primary invasive lung cancer cases (ICD-O-3 topographical codes C340-C349 excluding morphological codes 9590-9992) [26] by census tract, age group at diagnosis, and sex were obtained from the Cancer Data Registry of Idaho (CDRI) for the period 2007-2016. CDRI data are geocoded to support statistical analysis at the census tract and coarser levels of geographic detail. To geocode case data, CDRI used the Automated Geospatial Geocoding Interface Environment (AGGIE) System, developed under a partnership between the North American Association of Central Cancer Registries, Texas A&M University, and the National Cancer Institute to provide a single, uniform geocoding platform for open use by U.S. cancer registries [27,28]. The reference data utilized by the AGGIE System are the highest quality available from commercial and publicly available sources. In this study of 8,734 invasive lung and bronchus cancer diagnosed among Idaho residents between 2007-2016, inclusive, 94.2% were geocoded to the level of parcel or house number and street, 5.2% were geocoded using 5-digit ZIP Code centroid, and 0.5% using street mid-point or intersection. Five (0.06%) lung cancer cases diagnosed among Idaho residents during this time period were geocoded based on city name, with unknown ZIP Code, and could not be included in the study because census tract could not be accurately assigned.

Population estimates

Annual population counts by census tract, age, and sex are not available from the U.S. Census Bureau Population Estimates Program [29], so we estimated them. Starting with population counts by census tract, age, and sex that are available for the 2010 decennial census and annual county-level population estimates from the U.S. National Center for Health Statistics for 2007-2016 [30,31], we used iterative proportional fitting to allocate county population counts to tract by age and sex. Because this approach does not allow for differential growth over time by census tract within county, we supplemented this approach with data from serial time series from U.S. Census Bureau American Community Survey 5-year estimates (table B01001) covering 2007-2011 through 2012-2016 [32]. For two broad age groups, 0-39, and 40+, we performed linear regression on the natural log of the American Community Survey census tract population estimates to estimate annual percent change (APC). The APC estimates were used to allow differential trajectories, by broad age group, of census tract populations within each county. The census 2010 tract proportions were used directly for 2007-2010 because there was less population growth during the U.S. Great Recession. For 2011-2016, we used the tract-level APC trajectories constrained by the county-level population totals. The resultant population estimates have the characteristics of summing to the annual county-level population estimates

from the U.S. National Center for Health Statistics, having the same tract-level proportions within county for 2007-2010, and allowing differential growth for 2011-2016.

Smoking prevalence

For determining the prevalence of smoking in Idaho counties, we used data collected by the Bureau of Vital Records and Health Statistics, Division of Public Health, Idaho Department of Health and Welfare, under a cooperative agreement with the Centers for Disease Control and Prevention. These Behavioral Risk Factor Surveillance System (BRFSS) data are collected annually using a telephone survey that employs random sampling methods to measure population prevalence of risk factors for the major causes of death [33]. County-level estimates of current smoking for the period 1997-2007 and ever smoking for the period 2011-2016 were calculated by CDRI. Current smokers were respondents who reported having smoked at least 100 cigarettes in their lifetime and smoked every day or some days at the time of the survey interview. Ever smokers were respondents who reported having smoked at least 100 cigarettes in their lifetime at the time of the survey interview. The aggregated annual dataset analysis weights were poststratified to population estimates by age group, sex, and county, beginning with the BRFSS weights (poststratified to 2007 population estimates for 1997-2007 dataset and 2016 population estimates for 2011-2016 dataset). A minimum of 50 respondents was required to generate county-level statistics. As such, values for Camas and Clark Counties were imputed using the values for their respective Idaho public health district (South Central Public Health District and Eastern Idaho Public Health District, respectively).

Residential radon exposure

The Idaho Division of Public Health's Environmental Health Program provided residential radon test results by ZIP Code covering the period 1990 through February 2017 [34]. The results include only areas in Idaho that have been tested for radon and although the measurements are accurate, the results may not represent the distribution of radon levels throughout the state where low number of houses have been tested. We used a crosswalk between ZIP Code and census tract to be able to conduct the analysis at the census tract level [35]. The measure used in the statistical analysis was the proportion of test results greater than or equal to 4.0 picocuries per liter (pCi/L) of air in each census tract. At residential concentrations of 4.0 pCi/L or greater, the EPA recommends corrective actions to reduce radon exposure.

Particulate air pollution exposure

County estimates of fine particulate matter (PM 2.5 $\mu g/m^3$; air pollutants with an aerodynamic diameter less than 2.5 micrometers) in the outdoor air for the period 2001-2012 were obtained from the Centers for Disease Control and Prevention's Environmental Public Health Tracking Network [36]. We used annual average ambient concentrations of PM 2.5 based on monitors and modeling. For ecological measures, there is a balance between geographic specificity and accuracy. Estimates of fine particulate outdoor air pollution (PM 2.5) at the census tract level are available from the U.S. Environmental Protection Agency (EPA) via the EJSCREEN tool [37]. However, there is considerable uncertainty in the estimates for small areas and EJSCREEN is not used by EPA staff to quantify specific risk values for geographical areas. For these reasons, we opted to use the county-level estimates of fine particulate outdoor air pollution.

Socioeconomic position (SEP)

The percentage of persons residing in each census tract with incomes below poverty was categorized as < 5%, 5%-< 10%, 10%-< 20%, and ≥ 20% using estimates from the U.S. Census Bureau's American Community Survey for the period 2011-2015.

Statistical methods

All statistical analyses were conducted using SAS (version 9.4; Cary, NC) and SaTScan (version 9.4.4) [38]. We used SAS and SaTScan in combination for six sets of analyses. Poisson multilevel mixed models using SAS PROC GLIMMIX were fitted to model the incidence counts of lung cancer (see Formula 1). The data had a hierarchical multilevel structure with the lowest level consisting of age group by sex by census tract analysis cells. Analysis cells were nested within census tracts, which were nested within counties. Age group, sex, and their interaction were modeled as categorical variables. The age groups used were 0-39 (one combined group), 5-year age groups for 40-84, and 85+. All other measures were modeled as continuous variables. SAS PROC GIMMIX by default standardizes continuous variables to reduce collinearity by centering and scaling during estimation, but results are reported in terms of the original versions of the data. We conducted sensitivity analyses using categorical quantiles of the continuous variables, but found inferior fit using the pseudo-Akaike information criterion or increases in census tract covariance parameter estimates; these results are not reported. The natural logs of the corresponding census tract population estimates by age group and sex were used as offsets in the models. Census tract was included as a random effect in the models. Conditional on the census tract random effects and the fixed effects at the tract, county, and analysis cell levels, the observed cancer case counts were assumed to be independent Poisson variables. The models were optimized using the Newton-Raphson technique with ridging, starting from generalized linear model estimates.

Formula 1. Poisson multilevel mixed model of lung cancer incidence.

$\log(\mu_{ij}) = \log(P_{ij}) + \boldsymbol{\beta}\mathbf{X}_{ij} + \Upsilon_j$
$L_{ij} \sim \text{Poisson}(\mu_{ij})$
$\Upsilon_j \sim \text{Normal}(0, \delta^2)$
where: L_{ij} = count of lung cancer incident cases for i^{th} age by sex group, j^{th} census tract
μ_{ij}= expected number of lung cancer cases for i^{th} age by sex group, j^{th} census tract
P_{ij} = person years for i^{th} age by sex group, j^{th} census tract; offset for model
$\mathbf{X}_{ij}$ = column matrix of fixed effect covariates for age by sex group or census tract
$\boldsymbol{\beta}$ = row matrix of fixed effect regression coefficients
Υ_j = random effect intercept for j^{th} census tract

Expected case counts were obtained for each census tract, but excluded the census tract random effects in order to be used in the SaTScan analyses [23]. Likewise, covariance parameters to account for spatial autocorrelation, the extent to which data values for nearby census tracts are similar or dissimilar, were explicitly excluded from the models. The aim of the modeling process was not to develop the most parsimonious model, but rather to explore the impact of a multitude of area-based risk factors on the variability of census tract-level lung cancer incidence rates.

- The first model contained age group, sex, and their interaction.
- The second model added the two estimates of smoking prevalence, current smoking for 1997-2007 and ever smoked for 2011-2016.
- The third model added the residential radon measure and the interaction between radon and ever smoked.
- The fourth model added the fine particulate outdoor air pollution measure.
- The fifth model added elevation.
- The sixth model added census tract poverty category.

For each set of model results, discrete Poisson scan statistics were calculated using SaTScan to identify spatial clusters of lung cancer incidence at the census tract level [38]. The case file contained information on the observed total count of lung cancer cases for each census tract for 2007-2016, and was identical for each of the six runs. Latitude and longitude coordinates for each census tract were based on the population-weighted centroid [39]. The elevation in meters above mean sea level for the population-weighted centroids were obtained from the U.S. Geological Survey National Map – Elevation Point Query Service [40]. The population files contained information on the expected total counts of lung cancer cases for each census tract based on the six SAS PROC GLIMMIX models. Using a circular window, we scanned for both high and low rates of lung cancer incidence. The maximum allowable spatial cluster size was 50% of the population at risk. We used 999 Monte Carlo replications per SaTScan run, and p-values were calculated using software program defaults which include sequential Monte Carlo for large p-values and Gumbel approximations for very small p-values [41,42]. Using program parameters, we directed SaTScan to search for the most likely cluster using the likelihood function and for secondary clusters. Secondary clusters were selected hierarchically based on the most likely cluster for each centroid, with no geographic overlap with more likely clusters, and using the Gini index to optimize the reported secondary clusters [43]. Secondary clusters were evaluated without adjusting for more likely clusters. For each model, we report on clusters defined by p-values < 0.05.

Geographic data

Choropleth maps to depict standardized lung cancer incidence ratios by census tract (observed divided by expected counts for each model) and the location and size of clusters were created using QGIS software version 3.2 [44]. Census tract shapefiles are based on the U.S. Census Bureau's 2010 TIGER/Line files [45].

Results

Between 2007 and 2016, there were 8,734 incident primary lung cancer cases diagnosed among Idaho residents with address information sufficient to support analysis by census tract. Incidence rates increased with age and were higher for females from ages 40-54, and higher for males at older ages (Table 1). Among the 298 census tracts analyzed (2010 U.S. Census geography), the number of incident lung cancer cases diagnosed over the 10-year study period ranged from 0 to 93, with mean = 29.3 (Table 2). Age-standardized lung cancer incidence rates by census tract ranged from 0 to 108.3 cases per 100,000 population per year (2000 U.S. standard population with 19 age groups; Census P25-1130 standard). Idaho is a geographically large U.S. state, with area 216,440 km^2 and the 7^{th} least densely populated of the 50 U.S. states [46,47]. Locations of census tract population-weighted centroids span 42.1 to 48.8 degrees latitude, -117.0 to -111.1 degrees longitude, and elevations of the centroids range from 225.4 to 2311.7 meters. The proportion of residential radon sample results ≥ 4.0 pCi/L spanned 0.9% to 83.3% by census tract. Current smoking prevalence, measured using 1997-2007 aggregated BRFSS data, ranged 4.1% to 26.7% by county. Ever smoker prevalence, measured using 2011-2016 aggregated BRFSS data, ranged 9.1% to 56.6% by county. Fine particulate outdoor air pollution, measured by PM 2.5 concentration, ranged from 6.2 to 9.7 $\mu g/m^3$ by county. About four percent (4.4%) of Idaho's population resided in census tracts in the lowest poverty category (<5% below poverty) and 22.4% resided in census tracts in the highest poverty category (20%+ below poverty); a plurality (52.3%) resided in census tracts with 10-19.9% of the population below poverty.

Table 3 shows summaries of the results from SAS PROC GLIMMIX Poisson multilevel mixed models, including F statistics and p-values for the age group, sex, and age group by sex interaction categorical effects; parameter estimates (exponentiated beta, interpretable as incidence rate ratio for one-unit change on scale of data) and p-values for continuous variables; census tract covariance parameter estimates; and counts of statistically significant SaTScan clusters. Table 4 shows characteristics of the spatial clusters identified using SaTScan,

including the size of the circular window, the number of census tracts involved, the observed and expected numbers of lung cancer cases, the relative risk, and the p-value. Figures A1-A6 depict choropleth maps of standardized incidence ratios by census tract (observed divided by expected counts for each model) and the location and size of clusters.

Model 1

Results from the SAS PROC GLIMMIX Poisson multilevel mixed model showed significant age group, sex, and age group by sex interaction effects, which are consistent with the crude lung cancer incidence rates in Table 1. Accounting for population sizes by age group and sex, but not for additional ecological risk factors, SaTScan identified thirteen statistically significant clusters of census tracts: seven with higher incidence rates and six with lower rates (Table 4; Figure A1). The overall pattern of lung cancer incidence in Idaho includes a large area centered in southeast Idaho (97 of 298 tracts) with lower rates and a large area centered in west-central Idaho (132 tracts) with higher rates. There were six additional clusters of high rates and five additional clusters of low rates. Among these, all but one cluster of high rates had its center located within one of the large clusters.

Model 2

Because tobacco smoking is responsible for the majority of lung cancers, we searched for spatial clusters of high or low lung cancer incidence rates at the census tract level after adjusting for county-level smoking prevalence. Model 2 added continuous effects for county-level prevalence of current smoking (1997-2007) and ever smoker (2011-2016) to the age group, sex, and age group by sex interaction effects. The F statistics and p-values for the age group, sex, and age group by sex interaction effects were similar for all models, 1-6. Model 2 showed higher rates of lung cancer incidence at the census tract level, adjusting for age group and sex, with higher county-level prevalence of both the current smoking and ever smoker measures, which were independently statistically significant (Table 3). We calculated the percent change in the census tract covariance parameter, compared to Model 1, to approximate the degree to which adding area-level risk factor estimates decreased inter-tract variability. The inter-tract variability in lung cancer incidence rates was reduced by 54% by adjusting for county-level smoking prevalence. In other words, more than half of the differences in census tract-level lung cancer incidence were explained by county-level smoking prevalence. Accounting for population sizes by age group and sex, and for county-level smoking prevalence, SaTScan identified seven statistically significant clusters of census tracts: three with higher incidence rates and four with lower rates (Table 4; Figure A2). The most likely cluster included 28 census tracts in the Boise metropolitan area (RR=1.37). A large secondary cluster with lower rates of lung cancer centered in southeast Idaho (106 tracts) remained. There were two additional clusters of high rates, one that involved part of the Boise metropolitan area and one that involved part of the Lewiston metropolitan area. Two of the clusters of low rates had their centers located within the large cluster of low rates. The other cluster of low rates involved one census tract in northern Idaho (RR=0.27).

Model 3

Because residential radon exposure and the interaction between radon and smoking have been associated with 3-20% of lung cancer cases in developed countries, we searched for spatial clusters of high or low lung cancer incidence rates at the census tract level after adjusting for county-level smoking prevalence and radon. Model 3 added continuous effects for census-tract level residential radon and the interaction between tract-level radon and county-level prevalence of ever smoking to the effects used in Model 2. Model 3 showed higher rates of lung cancer incidence at the census tract level, adjusting for age group and sex, with higher values of the interaction between radon and ever smoker prevalence (Table 3). The point estimates for the constituent ever smoker prevalence and radon effects suggest they are protective of lung cancer; this is not the case, as they must be interpreted in the context of the interaction term. Although the term for ever smoker prevalence

became non-significant in this model, it was retained because it is a component of the interaction term [48]. Compared to Model 2, the magnitude of the effect for current smoking prevalence was reduced when radon and the radon by ever smoker interaction terms were included. Compared to Model 1, the inter-tract variability in lung cancer incidence rates was reduced 60% by adjusting for county-level smoking prevalence and tract-level radon. Accounting for population sizes by age group and sex, and for county-level smoking prevalence and tract-level radon, SaTScan identified seven statistically significant clusters of census tracts: two with higher incidence rates and five with lower rates (Table 4; Figure A3). The most likely cluster included 22 census tracts in the Boise metropolitan area (RR=1.34). A large secondary cluster with lower rates of lung cancer centered in southeast Idaho (78 tracts) remained. There was one additional cluster of high rates, a large cluster (30 census tracts) that included part of the Lewiston metropolitan area. Two of the clusters of low rates intersected with the large cluster of low rates. The same census tract in northern Idaho identified in the Model 2 SaTScan analysis as a cluster of low rates remained in this analysis (RR=0.28). An additional cluster of low rates emerged in this analysis that included eight census tracts in the Boise metropolitan area (RR=0.74).

Model 4

Model 4 added the continuous effect of a fine particulate outdoor air pollution measure, county-level PM 2.5 concentration, as 5-13% of lung cancer incidence in the developed world has been attributed to particulate air pollution. Model 4 showed significantly higher rates of lung cancer incidence at the census tract level with higher PM 2.5 concentration, adjusting for age group, sex, tract-level radon, and county-level smoking (Table 3). For a 1-unit increase in PM 2.5 $\mu g/m^3$, the incidence of lung cancer increased about 9%. Compared with Model 3, the addition of PM 2.5 to the model lessened the effect of the radon by ever smoker interaction and its constituent main effects but had little impact on the parameter estimate for county-level current smoker prevalence. Compared to Model 1, the inter-tract variability in lung cancer incidence rates was reduced 63% by adjusting for county-level smoking prevalence, tract-level radon, and county-level PM 2.5 concentration. Using the estimated tract-level counts from Model 4, SaTScan identified five statistically significant clusters of census tracts: two with higher incidence rates and three with lower rates (Table 4; Figure A4). The most likely cluster included 22 census tracts in the Boise metropolitan area (RR=1.28). The large secondary cluster with lower rates of lung cancer in southeast Idaho was centered in a different location and was greatly diminished in size, from 78 to 18 tracts. The large cluster that included part of the Lewiston metropolitan area remained (29 census tracts, RR=1.19). The same census tract in northern Idaho identified in the Models 2 and 3 SaTScan analyses as a cluster of low rates remained in this analysis (RR=0.26). The cluster of low rates in the Boise metropolitan area that emerged from Model 3 remained in this analysis (12 census tracts, RR=0.77).

Model 5

Model 5 added the continuous effect of elevation, measured as the elevation in meters above mean sea level for the population-weighted census tract centroids. Model 5 showed significantly higher lung cancer incidence rates with lower elevation, adjusting for age group, sex, tract-level radon, and county-level smoking and fine particulate outdoor air pollution (Table 3). Compared with Model 4, the addition of elevation to the model lessened the effect of PM 2.5 concentration and more modestly changed the parameter estimates for the radon and smoking prevalence measures. Compared to Model 1, the inter-tract variability in lung cancer incidence rates was reduced 65% by adjusting for tract-level radon and elevation and county-level smoking prevalence and PM 2.5 concentration. Using the estimated tract-level counts from Model 5, SaTScan identified three statistically significant clusters of census tracts: one with higher incidence rates and two with lower rates (Table 4; Figure A5). The most likely cluster included 22 census tracts in the Boise metropolitan area (RR=1.28). The same census tract in northern Idaho identified in the Models 2-4 SaTScan analyses as a cluster of low rates remained in this analysis (RR=0.26), as did the cluster of low rates in the Boise metropolitan area that emerged from Model 3 (12 census tracts, RR=0.77).

Model 6

Model 6 added the categorical effects of census tract poverty to Model 5. Model 6 showed significantly higher lung cancer incidence rates with higher county-level smoking, higher values of the interaction between radon and ever smoker prevalence, lower elevation, and census tract poverty category (Table 3). Compared to Model 1, the inter-tract variability in lung cancer incidence rates was reduced 77% by adjusting for the area-level covariates. Using the estimated tract-level counts from Model 6, SaTScan identified two statistically significant clusters of census tracts: one with higher incidence rates and one with lower rates (Table 4; Figure A6). The most likely cluster included 28 census tracts in the Boise metropolitan area (RR=1.20). The same census tract in northern Idaho identified in the Models 2-5 SaTScan analyses as a cluster of low rates remained in this analysis (RR=0.26). The cluster of low rates in the Boise metropolitan area that emerged from Model 3 was no longer significant after adjusting for census tract poverty.

Discussion

Although lung cancer incidence and mortality rates are on the decline, lung cancer continues to be the leading cause of cancer mortality among men and women in Idaho and the U.S. The combination of high incidence, high mortality, and low survival make lung cancer a vital public health issue. In this study, we used Poisson multilevel mixed models that included individual age group and sex and measurements of census tract-level elevation, residential radon, and SEP, and county-level smoking prevalence and fine particulate outdoor air pollution to understand spatial patterns of lung cancer incidence. At the census tract level, we found county smoking prevalence to be the leading driver of lung cancer incidence rates. Radon, the interaction between radon and smoking, air quality, elevation, and SEP further explained variation in census tract-level lung cancer incidence. There were fewer clusters after adjusting for area-level risk factors than with a model that adjusted only for age group and sex, signifying that the area-level risk factors are associated with the clusters in the simpler model. The clusters remaining after adjusting for the area-level risk factors show additional unexplained variation in lung cancer incidence that may warrant further study.

The relationships between area-level risk factors and lung cancer incidence in this study are consistent with other work. County-level smoking prevalence has previously been associated with higher rates of lung cancer in the U.S. [49]. Adjusting for smoking and socioeconomic status, a Korean study found a significant increase in lung cancer incidence among males with higher residential radon, and a similar point estimate among females [50]. A study of residential radon and cancer mortality in a region of Spain that adjusted for SEP, arsenic, and elevation, but not smoking, found a twofold increase in radon exposure was associated with a 9% increased risk of lung cancer mortality among women. Because smoking rates are substantially higher among males in this region of Spain, it was postulated that smoking masked the effect of radon exposure among males [20]. Area-level smoking prevalence has been found to be associated with lung cancer clustering in other U.S. states [51].

Substantial geographic variation in lung cancer incidence that exists in Idaho by census tract, as shown in the model that adjusted only for age group and sex (Table 4; Figure A1), was explained by the model that also included smoking (Figure A2). This finding was anticipated because the southeastern area of Idaho has a higher percentage of members of the Church of Latter-day Saints (LDS), among whom smoking prevalence is low [52]. After adjusting for smoking prevalence, there was less geographic clustering of low lung cancer incidence rates in southeastern Idaho.

The geology of Idaho, which includes areas with different ages of volcanic, metamorphic, and sedimentary rocks, contributes to geographic differences in soil radon and the majority of Idaho's land mass is characterized by the U.S. EPA as Radon Zone 1, the highest risk category for residential radon [11]. Including an interaction term for radon by smoking prevalence was important in this study. Because these risk factors are known to be synergistic [10], it is recommended that future studies of lung cancer geographic patterns using area-based measures include both main effects and interactions of radon and smoking prevalence.

Other studies have found a similar inverse relationship between elevation and lung cancer incidence. Van Pelt postulated the reason why elevation may be associated with lung cancer incidence is due to "the carcinogenic effect of higher absolute oxygen concentration in the inspired air at lower elevations" [18]. A Swiss study recognized a positive association of residential radon concentration with elevation [19]. Per the EPA map of Radon Zones, residential radon concentration is not positively associated with elevation at the county level in Idaho. We believe there is a different reason for the inverse relationship between elevation and lung cancer incidence. Idaho is topologically varied, with basin and range topography covering much of the southern portion of the state and forested mountains and deep valleys in the remainder. Especially during winter months, high pressure weather results in temperature inversions that concentrate cold, moist air and pollutants in low elevation valleys. Elevation could thus be a proxy measure for air pollution, picking up additional tract-level variation beyond the county-level PM 2.5 measure, and may be why the PM 2.5 measure became non-significant in the models that also included elevation. This finding may not be replicated in studies of other areas with different topography, such as less mountainous terrain or that are bordered by ocean.

Using the spatial scan statistic, the actual sizes of clusters are approximate, so it is not surprising that clusters that persisted from model to model may have changed slightly in extent [38]. The large clusters identified in Model 1 were no longer seen in later models. For example, the large cluster of low lung cancer incidence in southeast Idaho was not present in Model 3, which adjusted for age group, sex, tract-level radon, and smoking prevalence.

Findings from this study can be used to better target cancer control activities. We used six different models because results of intermediate models show differences in clusters and effects of covariates and lend insight into lung cancer incidence patterns attributable to different risk factors. For example, differences between the significant geographic clusters detected using data from Model 1, which included age group and sex, and Model 2, which also included smoking prevalence, can be used to target smoking cessation programmatic activity. Differences between the significant geographic clusters detected using data from Model 3 versus Models 4-6 suggest areas to focus on for outdoor air quality. Adding a SEP measure in Model 6 removed one significant cluster of low rates of lung cancer incidence, likely because inter-tract variability in smoking was better accounted for; lower SEP is associated with higher smoking prevalence in Idaho [53]. Future studies may endeavor to estimate smoking prevalence at a smaller geographic scale, such as postal code.

Using the expected census tract lung cancer case counts from Model 6, there was a cluster of high incidence rates identified in the Boise metropolitan area involving 28 census tracts (RR=1.20) and a cluster involving one tract with a low incidence rate in northern Idaho (RR=0.31). A potential reason for the cluster of high rates in part of the Boise metropolitan area may be air pollution. The fine particulate outdoor air pollution measure was county-based and does not differentiate among particulate sources such as wood fire smoke, wildfire smoke, diesel exhaust, industrial releases, and fugitive dust from construction, which may have different carcinogenic potential. The high-rate lung cancer cluster in this part of the Boise metropolitan area may be related to air pollution from transportation and industrial sources, which can be higher in metropolitan areas [54]. Near the center of this cluster is a spur rail line serving heavy industry and tank farms for petroleum

products. This cluster is also proximate to major highways. A study of local clustering of lung cancer in an area of Michigan found higher incidence to be related to proximity to point source pollution and major highways [55]. Further analysis of local air quality data and air pollution sources may better explain factors driving this cluster.

Major strengths of our study were the use of high quality geocoded cancer registry data as the source for counts of lung cancer incidence by census tract, age group, and sex, and a comprehensive set of ecological risk factors. Measures of lifestyle factors, air pollution, and other potential risk factors were at the census tract and county levels, not individual-level risk factors. Nonetheless, we found these ecological measures effective in reducing inter-tract variability and the number and extent of identified clusters. Individual-level measures of these risk factors are not available in the cancer registry database or through linkages. The area-level relationships we found between risk factors and lung cancer incidence may differ at the individual level. However, the ecological study design may be more germane to public health interventions at the community level.

The fine particulate outdoor air pollution measure was aggregated over 2001-2012 and had little variation, ranging from 6.2 to 9.7 PM 2.5 $\mu g/m^3$ by county. These concentrations are below the EPA's National Ambient Air Quality Standard of 12.0 $\mu g/m^3$ for annual average PM 2.5 [56]. However, large temporal differences in air pollution in some airsheds, caused by atmospheric conditions (temperature inversions) and wildfires, for example, were masked by virtue of averaging over a long time series. This may be one of the reasons that the air quality measure was non-significant in the models that also included elevation. The fact that the fine particulate outdoor air pollution values were based on monitors and modeling likely limited the variability.

Lung cancer may be diagnosed decades after exposure to carcinogens from smoking and other sources, and there was substantial immigration into Idaho prior to and during the study period. Some people diagnosed with lung cancer as residents of Idaho may have been exposed while living elsewhere and brought their constellations of exposures with them. Misclassification of exposures may have biased the results.

Estimates of area-level risk factors included different years of aggregated data and the lag periods between the risk factors and cancer incidence may not be biologically plausible at the individual level. We aggregated a long time series of both outcome and covariate measures in order to minimize variability due to small numbers of lung cancer cases and sampling errors in covariates. In particular, we used a long time series for the radon data, 1990 through February 2017, under the assumption that the natural release of radon varies little over time. We aggregated 10 years of incidence data in order to stabilize the census tract-specific lung cancer incidence rates. This strategy prevented us from investigating temporal or space-time clusters. In addition to the risk factors included in this study, there are additional risk factors for lung cancer, such as occupational exposures, arsenic in drinking water, personal or family history, and previous radiation therapy to the lungs. Recently, certain gene deletions have been shown to increase the risk of lung cancer, and may modulate the risk of lung cancer from radon [57]. Data on these risk factors were not available to us for inclusion in this study.

Conclusion

There are large geographic differences in lung cancer incidence within Idaho, including evidence of large and small clusters of high and low rates. Most of the differences in census tract-level lung cancer incidence in Idaho were explained by area-level measurements of known risk factors, in particular county-level smoking prevalence. After adjusting for ecological covariates, there were fewer and smaller clusters of lung cancer incidence. We found differences in clusters and effects of covariates from model to model that lend insight into lung cancer incidence patterns attributable to different ecological risk factors.

Compared to similar studies of lung cancer incidence, our study utilized a more comprehensive set of ecological risk factors. We postulate a novel explanation for the finding of increased lung cancer with lower elevation related to the concentration of fine particulate outdoor air pollutants under certain atmospheric and geographic conditions.

While smoking rates have declined over time, smoking prevalence in Idaho was 14.5% among adults in 2016, suggesting that the burden of lung cancer will be an important public health topic for decades to come, and that further efforts to decrease smoking are still needed [58]. The findings of this study may be used to identify geographic areas for cancer control initiatives by state and local health agencies and assist citizens and policymakers in understanding lung cancer spatial patterns in Idaho.

Table 1. Age and sex distribution of lung cancer cases and population in Idaho, 2007-2016.

	Males					Females				
Age Group (years)	Cases	Person-Years	Rate	Confidence Interval		Cases	Person-Years	Rate	Confidence Interval	
00-39	20	4568554	0.4	0.3	0.7	19	4366353	0.4	0.3	0.7
40-44	22	488173	4.5	2.8	6.8	32	477978	6.7	4.6	9.5
45-49	64	499362	12.8	9.9	16.4	81	498724	16.2	12.9	20.2
50-54	193	514212	37.5	32.4	43.2	203	524508	38.7	33.6	44.4
55-59	344	491320	70.0	62.8	77.8	291	504366	57.7	51.3	64.7
60-64	554	434248	127.6	117.2	138.7	516	443891	116.2	106.4	126.7
65-69	753	348834	215.9	200.7	231.8	724	357503	202.5	188.0	217.8
70-74	838	251744	332.9	310.7	356.2	742	263791	281.3	261.4	302.3
75-79	782	174095	449.2	418.2	481.8	646	195067	331.2	306.1	357.7
80-84	553	115865	477.3	438.3	518.8	491	148786	330.0	301.5	360.5
85+	430	94636	454.4	412.4	499.4	436	164429	265.2	240.9	291.3

Table 2. Summary statistics of census tract and county area-based measures.

Idaho census tracts (N=298)	Mean	Minimum	Median	Maximum
Lung cancer cases, 2007-2016	29.3	0.0	26.0	93.0
Person-years, 2007-2016	53444.4	420.0	46747.0	221996.0
Age-standardized lung cancer incidence rate, 2007-2016	51.2	0.0	51.7	108.3
Latitude of population-weighted centroid (NAD 1983)	44.3	42.1	43.6	48.8
Longitude of population-weighted centroid (NAD 1983)	-115.1	-117.0	-116.2	-111.1
Elevation (meters above mean sea level)	1042.1	225.4	863.4	2311.7
Radon, percent of samples in tract ≥ 4.0 pCi/L	32.1	0.9	28.5	83.3
	<5%	5-<10%	10-<20%	20%+
Percent of population by census tract poverty category	4.4%	20.8%	52.3%	22.4%

Idaho counties (N=44)	Mean	Minimum	Median	Maximum
Current smoking prevalence, 1997-2007	19.8	4.1	20.2	26.7
Ever smoker prevalence, 2011-2016	40.5	9.1	40.7	56.6
PM2.5 μg/m³	8.0	6.2	8.0	9.7

Table 3. Summary of SAS PROC GLIMMIX Poisson multilevel mixed models and discrete Poisson scan statistics calculated using SaTScan.

	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
Categorical Effects	F Value	Pr > F	F Value	Pr > F	F Value	Pr > F	F Value	Pr > F	F Value	Pr > F	F Value	Pr > F
Age Group	679.43	<.0001	678.26	<.0001	679.61	<.0001	680.31	<.0001	679.85	<.0001	680.08	<.0001
Sex	5.30	0.0214	5.09	0.0241	5.12	0.0236	5.18	0.0229	5.20	0.0226	5.22	0.0223
Age Group * Sex	6.62	<.0001	6.52	<.0001	6.54	<.0001	6.55	<.0001	6.54	<.0001	6.61	<.0001
Census Tract Poverty											17.27	<.0001
Continuous Effects	exp(Beta)	Pr > F	exp(Beta)	Pr > F	exp(Beta)	Pr > F	exp(Beta)	Pr > F	exp(Beta)	Pr > F	exp(Beta)	Pr > F
Current Smoker (1997-2007)			221.56	<.0001	110.79	<.0001	115.58	<.0001	93.03	<.0001	72.67	<.0001
Ever Smoker (2011-2016)			2.68	0.0248	0.29	0.1267	0.91	0.9165	0.40	0.2935	0.41	0.2581
Radon (percent ≥ 4 pCi/L)					0.06	<.0001	0.23	0.0387	0.20	0.0207	0.22	0.0194
Radon * Ever Smoker					388.93	<.0001	18.71	0.0812	38.76	0.0279	32.92	0.0210
Air Quality (PM 2.5 µg/m³)							1.09	0.0003	1.02	0.5513	1.04	0.1808
Elevation (meters)									0.99974	0.0010	0.99975	0.0004
Census Tract Covariance	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Parameter	0.09224	0.01152	0.04242	0.00697	0.03660	0.00639	0.03407	0.00610	0.03189	0.00589	0.02092	0.00494
Decrease from Model 1			54.0%		60.3%		63.1%		65.4%		77.3%	
SaTScan Clusters	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low
	7	6	3	4	2	5	2	3	1	2	1	1

Notes:

Predicted lung cancer case counts from models were used as expected “population” counts in SaTScan runs.

Exp(Beta) may be interpreted as the change in the adjusted incidence rate for a 1-unit change in the parameter for continuous effects. See Table 2 for ranges for the continuous effects.

Table 4. Statistically significant (p<0.05) high and low lung cancer incidence rate clusters identified by SaTScan, Idaho, 2007-2016.

Model 1. Age group, sex, age group by sex.

Cluster	Radius (km)	# Census tracts in cluster	Observed Cases	Expected Cases	Relative Risk	Gini Cluster	P-value
1	174.9	97	1550	2275.3	0.61	FALSE	<.0001
2	257.6	132	4879	4317.5	1.29	FALSE	<.0001
3	87.1	35	555	851.9	0.63	TRUE	<.0001
4	113.4	18	220	406.9	0.53	TRUE	<.0001
5	37.7	5	293	165.0	1.80	TRUE	<.0001
6	15.8	14	720	562.5	1.31	TRUE	<.0001
7	50.3	6	85	159.3	0.53	TRUE	<.0001
8	5.6	22	752	596.9	1.28	TRUE	<.0001
9	33.3	13	497	377.2	1.34	TRUE	<.0001
10	73.2	23	899	738.9	1.24	TRUE	<.0001
11	7.4	8	277	371.4	0.74	TRUE	0.0007
12	53.3	16	301	394.4	0.75	TRUE	0.0019
13	48.1	6	231	167.1	1.39	TRUE	0.0068

Model 2. Model 1 plus current smoking 1997-2007, ever smoker 2011-2016.

Cluster	Radius (km)	# Census tracts in cluster	Observed Cases	Expected Cases	Relative Risk	Gini Cluster	P-value
1	8.7	28	928	699.4	1.37	TRUE	<.0001
2	286.3	106	1776	2062.7	0.83	FALSE	<.0001
3	113.4	18	220	326.0	0.67	TRUE	<.0001
4	96.1	19	484	611.4	0.78	TRUE	0.0001
5	0.0	1	9	33.3	0.27	TRUE	0.0017
6	25.5	29	1053	918.2	1.17	TRUE	0.0100
7	38.2	11	293	223.6	1.32	TRUE	0.0140

Model 3. Model 2 plus radon, radon x ever smoker.

Cluster	Radius (km)	# Census tracts in cluster	Observed Cases	Expected Cases	Relative Risk	Gini Cluster	P-value
1	5.6	22	752	573.7	1.34	TRUE	<.0001
2	168.8	78	1578	1782.7	0.86	FALSE	0.0002
3	113.4	18	220	306.5	0.71	TRUE	0.0006
4	7.4	8	277	368.7	0.74	TRUE	0.0015
5	0.0	1	9	32.3	0.28	TRUE	0.0040
6	68.5	18	343	437.2	0.78	TRUE	0.0056
7	117.0	30	1064	927.8	1.17	TRUE	0.0130

(Table 4 continued)

Model 4. Model 3 plus air quality (PM 2.5 µg/m³).

Cluster	Radius (km)	# Census tracts in cluster	Observed Cases	Expected Cases	Relative Risk	Gini Cluster	P-value
1	5.6	22	752	598.2	1.28	TRUE	<.0001
2	8.8	12	473	606.2	0.77	TRUE	<.0001
3	0.0	1	9	34.6	0.26	TRUE	0.0010
4	113.1	29	999	856.4	1.19	TRUE	0.0021
5	113.4	18	220	299.2	0.73	TRUE	0.0038

Model 5. Model 4 plus elevation.

Cluster	Radius (km)	# Census tracts in cluster	Observed Cases	Expected Cases	Relative Risk	Gini Cluster	P-value
1	5.6	22	752	596.8	1.28	TRUE	<.0001
2	8.8	12	473	605.1	0.77	TRUE	<.0001
3	0.0	1	9	34.1	0.26	TRUE	0.0010

Model 6. Model 5 plus census tract poverty

Cluster	Radius (km)	# Census tracts in cluster	Observed Cases	Expected Cases	Relative Risk	Gini Cluster	P-value
1	8.7	28	928	789.4	1.20	TRUE	0.0014
2	0.0	1	9	28.9	0.31	TRUE	0.0300

Figures A1-A6. Maps of census tract incidence rate ratios (observed divided by expected lung cancer case counts from each SAS PROC GLIMMIX Poisson multilevel mixed model) overlaid by statistically significant clusters of high and low rates identified using discrete Poisson scan statistics, Idaho 2007-2016. Figure A1 shows results for Model 1, Figure A2 shows results from Model 2, etc. (see Table 3). The labels for the clusters refer to the cluster numbers in Table 4.

Figure A1

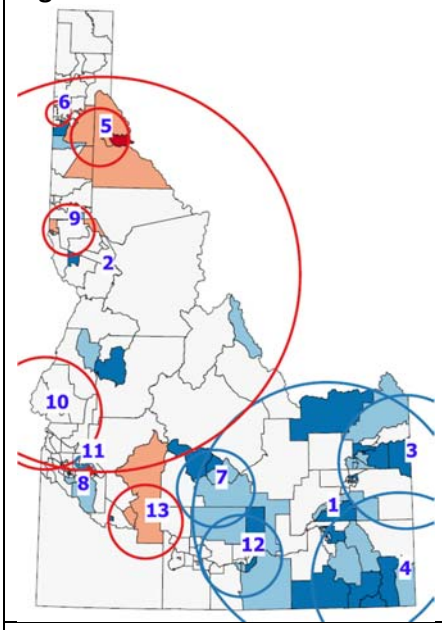


Figure A2

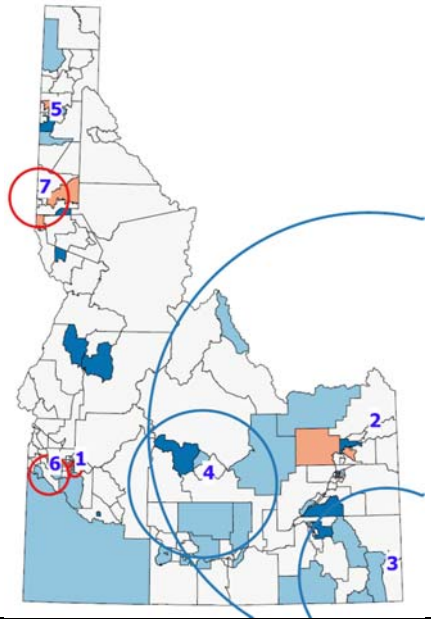


Figure A3

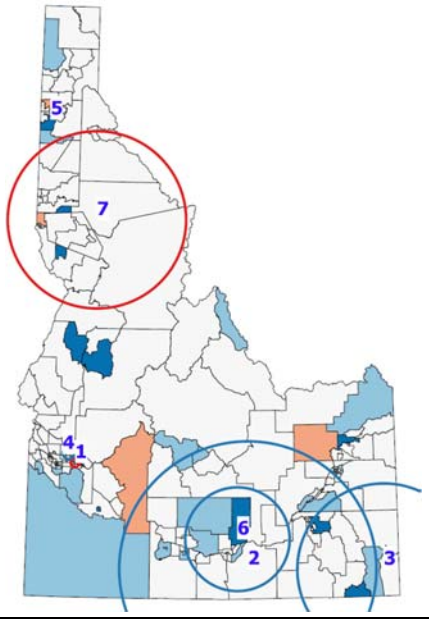


Figure A4

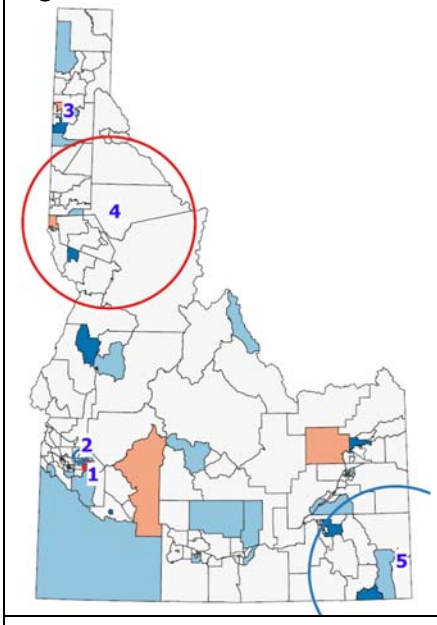


Figure A5

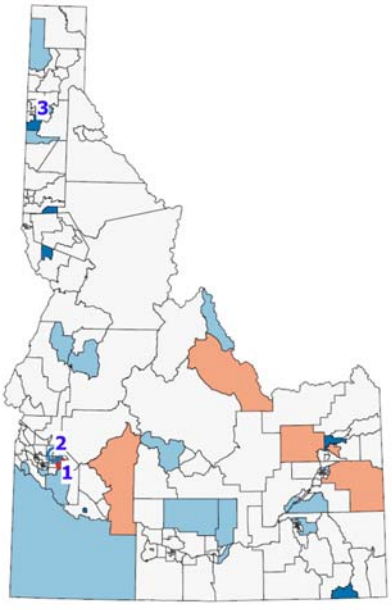
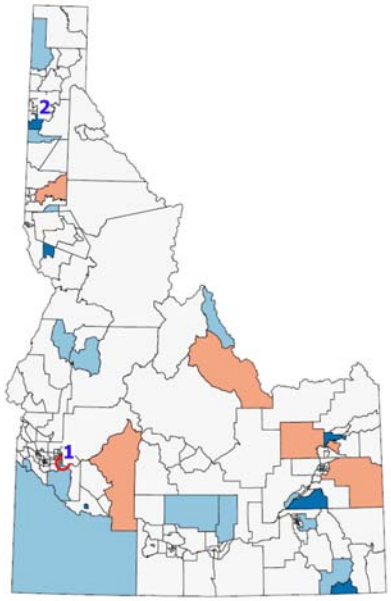


Figure A6



Legend

Census Tract Incidence Rate Ratio

- 0.00 - 0.49
- 0.49 - 0.67
- 0.67 - 1.50
- 1.50 - 2.00
- 2.00 - 2.45

Cluster Relative Risk (circles)

- 0.00 - 1.00
- 1.00 - 1.80

References

1. Idaho Cancer/Cluster Analysis Work Group. Investigation of Non-infectious Disease Clusters. 2014. http://www.idcancer.org/Documents/Resources/CAWG%20Investigation%20of%20Non-infectious%20Disease%20Clusters%202014.pdf. Accessed 20 March 2018.

2. Centers for Disease Control and Prevention. Investigating Suspected Cancer Clusters and Responding to Community Concerns: Guidelines from CDC and the Council of State and Territorial Epidemiologists. Morbidity and Mortality Weekly Report. 2013;62(RR08):1-14.

3. Bray F, Colombet M, Mery L, Piñeros M, Znaor A, Zanetti R and Ferlay J, editors.
Cancer Incidence in Five Continents, Vol. XI. International Agency for Research on Cancer. 2017. http://ci5.iarc.fr. Accessed 20 March 2018.

4. Johnson CJ, Carson SL. Cancer in Idaho, 2015. Boise: Cancer Data Registry of Idaho; 2017. https://www.idcancer.org/ContentFiles/AnnualReports/Cancer%20in%20Idaho%202015.pdf (page 6). Accessed 20 March 2018.

5. State Cancer Profiles. Quick Profiles: Idaho. https://statecancerprofiles.cancer.gov/quick-profiles/index.php?statename=idaho#t=4. Accessed 20 March 2018.

6. Meza R, Meernik C, Jeon J, Cote ML. Lung Cancer Incidence Trends by Gender, Race and Histology in the United States, 1973–2010. PLoS ONE 2015;10(3):e0121323. https://doi.org/10.1371/journal.pone.0121323. Accessed 20 March 2018.

7. U.S. Cancer Statistics Working Group. United States Cancer Statistics: 1999-2014 Incidence and Mortality Web-based Report. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute. 2017. https://nccd.cdc.gov/uscs. Accessed 20 March 2018.

8. U.S. Department of Health and Human Services. The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General, 2014. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. 2014. https://www.surgeongeneral.gov/library/reports/50-years-of-progress/index.html. Accessed 20 March 2018.

9. Gillmore GK, Phillips PS, Denman AR. The effects of geology and the impact of seasonal correction factors on indoor radon levels: A case study approach. J Environ Radioact. 2005;84:469-479.

10. U.S. Environmental Protection Agency. Health Risk of Radon. https://www.epa.gov/radon/health-risk-radon. Accessed 20 March 2018.

11. U.S. Environmental Protection Agency. EPA Map of Radon Zones.
https://www.epa.gov/sites/production/files/2015-07/documents/zonemapcolor.pdf. Accessed 20 March 2018.

12. López-Cima MF, García-Pérez J, Pérez-Gómez B, et al. Lung cancer risk and pollution in an industrial region of Northern Spain: a hospital-based case-control study. International Journal of Health Geographics. 2011;10:10. doi:10.1186/1476-072X-10-10.

13. Christian WJ, Huang B, Rinehart J, Hopenhayn C. Exploring geographic variation in lung cancer incidence in Kentucky using a spatial scan statistic: elevated risk in the Appalachian coal-mining region. Public Health Rep. 2011 Nov-Dec;126(6):789-96.

14. Hecht SS. Tobacco smoke carcinogens and lung cancer. J Natl Cancer Inst. 1999;91:1194–1210.

15. Kim S-H, Hwang WJ, Cho J-S, Kang DR. Attributable risk of lung cancer deaths due to indoor radon exposure. Annals of Occupational and Environmental Medicine. 2016;28:8. doi:10.1186/s40557-016-0093-4.

16. Chen H, Goldberg MS, Villeneuve PJ. A systematic review of the relation between long-term exposure to ambient air pollution and chronic diseases. Rev Environ Health. 2008;23(4):243–297.

17. Pope CA, Burnett RT, Turner MC, et al. Lung Cancer and Cardiovascular Disease Mortality Associated with Ambient Air Pollution and Cigarette Smoke: Shape of the Exposure–Response Relationships. Environmental Health Perspectives. 2011;119(11):1616-1621. doi:10.1289/ehp.1103639.

18. Van Pelt WR. Epidemiological associations among lung cancer, radon exposure and elevation above sea level--a reassessment of Cohen's county level radon study. Health Phys. 2003 Oct;85(4):397-403.

19. Kropat G, Bochud F, Jaboyedoff M, et al. Major influencing factors of indoor radon concentrations in Switzerland. J Environ Radioact. 2014;129:7–22. http://dx.doi.org/10.1016/j.jenvrad.2013.11.010.

20. López-Abente G, Núñez O, Fernández-Navarro P, Barros-Dios JM, Martín-Méndez I, Bel-Lan A, Locutura J, Quindós L, Sainz C, Ruano-Ravina A. Residential radon and cancer mortality in Galicia, Spain. Sci Total Environ. 2018 Jan 1;610-611:1125-1132. doi: 10.1016/j.scitotenv.2017.08.144. Epub 2017 Aug 30.

21. Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian SV, Carson R. Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter?: The Public Health Disparities Geocoding Project. Am J Epidemiol 2002; 156:471-482.

22. Boscoe FP, Johnson CJ, Sherman RL, Stinchcomb DG, Lin G, Henry KA. The Relationship Between Area Poverty Rate and Site-Specific Cancer Incidence in the United States. Cancer. 2014;120(14):2191-2198. doi:10.1002/cncr.28632.

23. Klassen AC, Kulldorff M, Curriero F. Geographical clustering of prostate cancer grade and stage at diagnosis, before and after adjustment for risk factors. Int J Health Geogr. 2005 Jan 13;4(1):1.

24. Bulka CM, Jones RM, Turyk ME, Stayner LT, Argos M. Arsenic in drinking water and prostate cancer in Illinois counties: An ecologic study. Environ Res. 2016 Jul;148:450-456. doi: 10.1016/j.envres.2016.04.030. Epub 2016 Apr 29.

25. Christian WJ, Huang B, Rinehart J, Hopenhayn C. Exploring geographic variation in lung cancer incidence in Kentucky using a spatial scan statistic: elevated risk in the Appalachian coal-mining region. Public Health Rep. 2011 Nov-Dec;126(6):789-96.

26. Fritz A, Percy C, Jack A, Shanmugaratnam K, Sobin L, Parkin D, Whelan S. International Classification of Diseases for Oncology. 3rd ed. Geneva, Switzerland: World Health Organization; 2000.

27. Goldberg DW, Kohler B, Kosary C. The Texas A&M, NAACCR, NCI Geocoding Service. 2017. http://geo.naaccr.org. Accessed 20 March 2018.

28. Goldberg DW, Cockburn MG. Improving geocode accuracy with candidate selection criteria. Transactions in GIS. 2010;14(S1):129-146.

29. United States Census Bureau Population Estimates Program. 2017. https://factfinder.census.gov/faces/nav/jsf/pages/programs.xhtml?program=pep. Accessed 20 March 2018.

30. National Center for Health Statistics. Bridged-race intercensal estimates of the resident population of the United States for July 1, 2000-July 1, 2009, by year, county, single-year of age (0, 1, 2, .., 85 years and over), bridged race, Hispanic origin, and sex. Prepared under a collaborative arrangement with the U.S. Census Bureau. https://www.cdc.gov/nchs/nvss/bridged_race.htm. Accessed 20 March 2018.

31. National Center for Health Statistics. Vintage 2016 postcensal estimates of the resident population of the United States (April 1, 2010, July 1, 2010-July 1, 2016), by year, county, single-year of age (0, 1, 2, .., 85 years and over), bridged race, Hispanic origin, and sex. Prepared under a collaborative arrangement with the U.S. Census Bureau. https://www.cdc.gov/nchs/nvss/bridged_race.htm. Accessed 20 March 2018.

32. United States Census Bureau. American Community Survey 5-Year Estimates, Table B01001. 2017. https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_16_5YR_B01001&prodType=table. Accessed 20 March 2018.

33. Idaho Department of Health and Welfare. Behavioral Risk Factor Surveillance System. http://www.healthandwelfare.idaho.gov/Health/VitalRecordsandHealthStatistics/HealthStatistics/BehavioralRiskFactorSurveillanceSystem/tabid/913/Default.aspx. Accessed 20 March 2018.

34. Idaho Department of Health and Welfare. Indoor Environment: Radon. http://healthandwelfare.idaho.gov/Health/EnvironmentalHealth/IndoorEnvironment/Radon/tabid/939/Default.aspx. Accessed 20 March 2018.

35. U.S. Department of Housing and Urban Development. HUD USPS ZIP Code Crosswalk Files. https://www.huduser.gov/portal/datasets/usps_crosswalk.html. Accessed 20 March 2018.

36. Centers for Disease Control and Prevention. Environmental Public Health Tracking Network. Annual PM 2.5 Level (Monitor + Modeled). https://ephtracking.cdc.gov. Accessed 20 March 2018.

37. United States Environmental Protection Agency. EJSCREEN: Environmental Justice Screening and Mapping Tool. https://www.epa.gov/ejscreen/what-ejscreen. Accessed 20 March 2018.

38. Kulldorff M. A spatial scan statistic. Communications in Statistics: Theory and Methods 1997;26:1481-1496.

39. United States Census Bureau. Geography: Centers of Population. https://www.census.gov/geo/reference/centersofpop.html. Accessed 20 March 2018.

40. United States Geological Survey. The National Map - Elevation Point Query Service. https://nationalmap.gov/epqs/. Accessed 20 March 2018.

41. Besag J, Clifford J. Sequential Monte Carlo p-values. Biometrika 1991;78:301-330.

42. Abrams A, Kleinman K, Kulldorff M. Gumbel based p-value approximations for spatial scan statistics. Int J Health Geogr. 2010 Dec 17;9:61. doi: 10.1186/1476-072X-9-61.

43. Han J, Zhu L, Kulldorff M, Hostovich S, Stinchcomb DG, Tatalovich Z, Lewis DR, Feuer EJ. Using Gini coefficient to determining optimal cluster reporting sizes for spatial scan statistics. Int J Health Geogr. 2016 Aug 3;15(1):27. doi: 10.1186/s12942-016-0056-6.

44. QGIS Development Team. QGIS Geographic Information System. Open Source Geospatial Foundation. 2018. https://qgis.org/en/site/. Accessed 3 August 2018.

45. United States Census Bureau. TIGER/Line Shapefiles and TIGER/Line Files. https://www.census.gov/geo/maps-data/data/tiger-line.html. Accessed 20 March 2018.

46. United States Census Bureau, Population Division. Annual Estimates of the Population for the United States, Regions, States, and Puerto Rico: April 1, 2010 to July 1, 2015. December 2015. https://www2.census.gov/programs-surveys/popest/datasets/2010-2015/state/asrh/. Accessed 20 March 2018.

47. United States Census Bureau. Geography: State Area Measurements and Internal Point Coordinates. 2010 United States Census. https://www.census.gov/geo/reference/state-area.html. Accessed 20 March 2018.

48. Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research: principles and quantitative methods. Belmont, California: Lifetime Learning Publications; 1982.

49. Lewis DR, Pickle LW, Zhu L. Recent Spatiotemporal Patterns of US Lung Cancer by Histologic Type. Front Public Health. 2017 May 19;5:82. doi: 10.3389/fpubh.2017.00082.

50. Ha M, Hwang SS, Kang S, Park NW, Chang BU, Kim Y. Geographical Correlations between Indoor Radon Concentration and Risks of Lung Cancer, Non-Hodgkin's Lymphoma, and Leukemia during 1999-2008 in Korea. Int J Environ Res Public Health. 2017 Mar 24;14(4). pii: E344. doi: 10.3390/ijerph14040344.

51. Dietz NA, Sherman R, Mackinnon J, Fleming L, Arheart KL, Wohler B, Lee DJ. Toward the identification of communities with increased tobacco-associated cancer burden: Application of spatial modeling techniques. J Carcinog. 2011;10:22. doi: 10.4103/1477-3163.85184. Epub 2011 Sep 21.

52. Merrill RM, Hilton SC, Daniels M. Impact of the LDS church's health doctrine on deaths from diseases and conditions associated with cigarette smoking. Ann Epidemiol. 2003 Nov;13(10):704-11.

53. Idaho Department of Health and Welfare. Idaho Health Behaviors, 2015. Results from the Behavioral Risk Factor Surveillance System. 2017. http://healthandwelfare.idaho.gov/Portals/0/Health/Statistics/BRFSS%20Reports/Idaho_BRFSS_Annual_Report_2015.pdf. Accessed 20 March 2018.

54. Cohen AJ, Pope CA 3rd. Lung cancer and air pollution. Environ Health Perspect. 1995;103 Suppl 8:219-24.

55. Guajardo OA, Oyana TJ. A critical assessment of geographic clusters of breast and lung cancer incidences among residents living near the Tittabawassee and Saginaw Rivers, Michigan, USA. J Environ Public Health. 2009;2009:316249. doi: 10.1155/2009/316249. Epub 2009 Nov 25.

56. United States Environmental Protection Agency. Particulate Matter (PM) Pollution: Fact Sheets and Additional information Regarding the 2012 Particulate Matter (PM) National Ambient Air Quality Standards (NAAQS). https://www.epa.gov/pm-pollution/fact-sheets-and-additional-information-regarding-2012-particulate-matter-pm-national. Accessed 20 March 2018.

57. Ruano-Ravina A, Pereyra MF, Castro MT, Pérez-Ríos M, Abal-Arca J, Barros-Dios JM. Genetic susceptibility, residential radon, and lung cancer in a radon prone area. J Thorac Oncol. 2014 Aug;9(8):1073-80. doi: 10.1097/JTO.0000000000000205.

58. United States Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Division of Population Health. BRFSS Prevalence & Trends Data [online]. 2015. https://www.cdc.gov/brfss/brfssprevalence/. Accessed 20 March 2018.