



Miscellaneous

Lung and stomach cancer associations with groundwater radon in North Carolina, USA

Kyle P Messier and Marc L Serre*

Department of Environmental Science and Engineering, Gillings School of Global Public Health, University of North Carolina, USA

*Corresponding author. Department of Environmental Sciences and Engineering, Gillings School of Global Public Health, University of North Carolina, 1303 Michael Hooker Research Center, Chapel Hill, NC 27599, USA. E-mail: marc_serre@unc.edu

Abstract

Background: The risk of indoor air radon for lung cancer is well studied, but the risks of groundwater radon for both lung and stomach cancer are much less studied, and with mixed results.

Methods: Geomasked and geocoded stomach and lung cancer cases in North Carolina from 1999 to 2009 were obtained from the North Carolina Central Cancer Registry. Models for the association with groundwater radon and multiple confounders were implemented at two scales: (i) an ecological model estimating cancer incidence rates at the census tract level; and (ii) a case-only logistic model estimating the odds that individual cancer cases are members of local cancer clusters.

Results: For the lung cancer incidence rate model, groundwater radon is associated with an incidence rate ratio of 1.03 [95% confidence interval (CI) = 1.01, 1.06] for every 100 Bq/l increase in census tract averaged concentration. For the cluster membership models, groundwater radon exposure results in an odds ratio for lung cancer of 1.13 (95% CI = 1.04, 1.23) and for stomach cancer of 1.24 (95% CI = 1.03, 1.49), which means groundwater radon, after controlling for multiple confounders and spatial auto-correlation, increases the odds that lung and stomach cancer cases are members of their respective cancer clusters.

Conclusion: Our study provides epidemiological evidence of a positive association between groundwater radon exposure and lung cancer incidence rates. The cluster membership model results find groundwater radon increases the odds that both lung and stomach cancer cases occur within their respective cancer clusters. The results corroborate previous biokinetic and mortality studies that groundwater radon is associated with increased risk for lung and stomach cancer.

Key words: Radon, groundwater, stomach cancer, lung cancer, generalized linear models, cluster analysis

Key Messages

- Modelled census tract average groundwater radon is associated with a positive increase in lung cancer incidence rates.
- To supplement the ecological scale analysis, we performed a two-stage analysis consisting of a cancer cluster analysis followed by logistic regression on cancer case membership within the cluster.
- Modelled address-level groundwater radon is associated with an increase in the probability of a lung cancer case occurring within a local lung cancer cluster.
- Modelled address-level groundwater radon is associated with an increase in the probability of a stomach cancer case occurring within a local stomach cancer cluster.

Introduction

Radon is a naturally occurring radioactive gas and human carcinogen found in the groundwater drinking supply and indoor air across the world. Countries with documented groundwater radon occurrence include the USA,^{1–3} Finland,⁴ Belgium,⁵ Italy⁶ and many other European countries.⁷ The carcinogenic risk associated with radon exposure is due to its radioactive decay and emission of high energy alpha decay particles (α -decay);^{8,9} thus when referring to radon, it is generally understood to be radon and its associated α -decay.

There is vast literature including multiple epidemiological analyses supporting the conclusion that exposures via inhalation of radon in indoor air lead to a significant increased risk of lung cancer morbidity in both never-smokers and smokers.^{7,10–14} The International Agency for Research on Cancer concluded there is sufficient evidence that radon and its progeny cause lung cancer in humans¹⁵. Ingestion of radon is also thought to be associated with lung cancer; however, the literature for the groundwater or drinking-water routes of exposure and lung cancer is limited to biokinetic models^{8,16} and one ecological epidemiology analysis of mortality.¹⁷ Although the source of contamination, subsurface geology containing radon or its parent chemicals, is often the same for indoor air and groundwater radon, their effect on lung cancer cannot be assumed to be the same because they have differing environmental levels (since air radon enters the home via infiltration through the soil and home structure, whereas groundwater enters through the private well), different routes of exposure (primarily inhalation for indoor air radon and ingestion for groundwater radon⁸), and are subject to different remediation measures (air versus water).

Stomach cancer is likely to be the second major cancer risk from radon exposure, after lung cancer;^{8,9,11} however, previous studies have looked at stomach cancer and radon with mixed results.¹¹ A case-cohort study of private well radon found a protective effect that was not statistically significant; however, it most likely suffered from a small

cohort ($n = 371$) and lack of confounder- controlling for unmeasured protective effects.⁴ A county-scale ecological analysis found a positive relationship between indoor air radon and stomach cancer mortality; however, the study did not report the number of subjects or the confidence intervals.¹⁸ Kendall and Smith¹¹ hypothesized that the mixed results of stomach cancer studies are because they lack a highly exposed cohort of sufficient sample size.

North Carolina contains geological features commonly associated with elevated radon and has many areas across the state with high concentration of radon in the groundwater.³ Additionally, Messier *et al.*³ recently provided a geologically based land use regression and Bayesian Maximum Entropy (LUR-BME) model to spatially estimate groundwater radon, affording the possibility to quantify exposure across the state at a fine resolution not previously possible with other models of groundwater radon exposure. Lastly, state-wide lung cancer incidence rates are higher than the national average for 2007–11 (72.7 versus 64.9 per 100 000 person-years) and near the national average for stomach cancer (6.7 versus 6.3 per 100 000 person-years).¹⁹

The objectives of our study are to: (i) provide an epidemiological analysis of groundwater radon exposure and lung cancer incidence; and (ii) conduct an epidemiological analysis of groundwater radon and stomach cancer incidence with a large and exposed cohort. To this end, we develop two types of models for lung and stomach cancer in North Carolina across an 11-year period. The first type of model examines associations at an ecological scale, investigating the association of groundwater radon exposure and lung and stomach cancer incidence rates by census tract. To expand upon the ecological-level model, we develop a two-stage cluster analysis and logistic regression framework that estimates the odds that cancer cases belong to cancer clusters, which allows for an assessment at the individual as opposed to ecological scale. This framework has been applied to evaluating the associations between H5N1 avian bird flu and environmental factors,^{20,21} amyotrophic

lateral sclerosis and lake water quality,²² and tuberculosis and aboriginal ancestry.²³

Results will be of interest to cancer researchers across disciplines including toxicologists and epidemiologists, federal and state agencies monitoring public health such as the Department of Health and Human Services, and to the general public in order to become better educated on their potential risks associated with groundwater radon exposure. Furthermore, the results will provide the relative risk estimate needed to calculate the sample size for a large case-control study of radon and cancer outcomes, which will be significantly more expensive and time-consuming than this study.

Methods

Study population

Geomasked address-level stomach and lung incident cancer cases in North Carolina from 1999 to 2009 were obtained from the North Carolina Central Cancer Registry (NCCCR) with a data use agreement. An Internal Review Board (IRB) assessment was obtained (UNC-IRB #12-1761) for human subjects; however, the only identifiable information is their location. Geomasked locations are moved slightly from true addresses using a donut geomask to protect privacy while preserving the sensitivity and specificity of detecting disease clusters.^{24,25} Attributes include race, age at diagnosis, gender (Table 1) and various notes including tobacco use history; however, those are reported in less than 10% of cases. Stages of cancer were also not included.

Exposure data

Groundwater radon concentration (Bq/L) exposure is estimated from Messier *et al.*,³ which are address-level estimates of groundwater radon concentration based on a land use regression and Bayesian Maximum Entropy (LUR-BME) geostatistical model with geologically based explanatory variables. The LUR-BME model provides address-level estimates of groundwater radon concentration for individual exposure assessment, and spatial averaging of estimates provides a precise assessment of observed census tract levels (Pearson correlation = 0.9; see [supplementary material](#) for more details on the LUR-BME exposure models and their validation, available as [Supplementary data at IJE online](#)).

Statistical analyses at multiple spatial scales

Associations between stomach and lung cancer are examined at two different spatial scales.

Table 1. Basic information for the study population. Lung and stomach cancer cases from 1999 to 2009 in North Carolina, USA

	Stomach cancer	Lung cancer
Male		
White		
Age < 65	814	10080
Age ≥ 65	1345	20065
Black		
Age < 65	423	3099
Age ≥ 65	457	3244
Other		
Age < 65	55	217
Age ≥ 65	34	219
Female		
White		
Age < 65	413	7663
Age ≥ 65	960	15083
Black		
Age < 65	236	1776
Age ≥ 65	401	2006
Other		
Age < 65	41	161
Age ≥ 65	39	191
Total	5 218	63 804

First, incidence rates are examined at the census tract level using a negative binomial generalized linear model (GLM) with standard NB2 parameterization^{26,27} (referred to hereinafter as the incidence rate model). The model of stomach or lung cancer counts, Y , is assumed to follow a negative binomial distribution such that $Y \sim NB2(\mu, \alpha)$, where μ is the mean and α is the negative binomial dispersion parameter. For the NB2 parameterization, the natural log is the link function and the exponential is the inverse-link; thus we model cancer counts as:

$$\ln(Y) = \beta_0 + \beta_1 Z_1 + \dots + \beta_n Z_n + \varepsilon + \text{offset} \quad (1)$$

where Y is the number of stomach or lung cancer counts in a given census tract over the 11-year study period, β_n are linear coefficients for the census tract predictor variables Z_n , ε is the error term, and *offset* is the population-year offset, which is the natural log of the census tract population times the duration of the study period (11 years) with a coefficient constrained to 1 resulting in an incidence rate interpretation of the model.

The predictor variables include the exposure Z_1 of interest (the census tract average of groundwater radon concentration, Bq/L), and known confounding variables, $Z_l, l > 1$, which include indoor air radon exposure, smoking prevalence,²⁸ public water supply status, residential tenure, age, gender and race. The indoor air radon census tract

geometric mean is estimated via *BMElib* with Gaussian soft data²⁹ based on of 6320 indoor air basement measurements across North Carolina, which approximates the census tract population geometric mean (Pearson correlation > 0.5; see [supplementary material](#), available as [Supplementary data at IJE online](#)). Using census tract geometric means reduces variability and has already been done in other studies of groundwater pollutants.³⁰ Here we use census tract geometric means of basement measurements to remove the variability due to individual house characteristics, which are unavailable for our cancer cases. It should be noted that although we control for air radon, the correlation between groundwater and air radon is weak (R-squared = 0.09) based on 238 log-transformed co-located groundwater and indoor air measurements, and therefore we do not expect that air radon is a strong confounder for groundwater radon in this study. More details on the estimation of indoor air radon and other confounding variables are available in the [supplementary material](#) (available as [Supplementary data at IJE online](#)). Incidence rate ratio (IRR), or the ratio of the probabilities of disease when a given predictor variable is increased by one unit, is obtained for each variable by exponentiating its coefficient ($IRR = e^\beta$). We create and compare models with increasing levels of control for confounding variables. First, a crude model or model with only groundwater radon is produced. Second, in the adjusted model we control for gender and age with dummy variables. Last, we control for all confounders including indoor air radon, smoking, public water supply, residential tenure, race, gender and age with a single full model. To utilize the address-level exposure information from the groundwater radon estimates,³ we conduct a logistic regression analysis on lung and stomach cancer cases that are assigned a 0/1 status based on their membership in a cluster^{20,22} (referred hereinafter as the cluster membership model). Cancer clusters are identified by calculating the Anselin Local Moran's I on normalized excess case counts:²²

$$c_i = (o_i - e_i)/e_i$$

where o_i is the number of observed cancer cases per census tract and e_i is the expected number of cases calculated as the North Carolina state average for the study period and gender- and age-adjusted for each census tract. These cancer clusters may represent geographical regions with unknown elevated risk factors. This approach allows address-level exposure information to be utilized in case-only studies and where a case-crossover study design is not sensible. To identify these risk factors, we assign each individual cancer cases with a 0/1 binary variable M indicating their membership in cancer clusters. We model the probability that a lung or stomach cancer is a member of a

cancer cluster using the logistic generalized linear model (GLM):

$$\text{logit}(M) = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n + \varepsilon \quad (2)$$

where $\text{logit}(M)$ is the logit link function that transforms the binary membership dependent variable M to the appropriate scale for estimation, β_n are linear coefficients for the individual predictor variables X_n and ε is the error term. Groundwater radon at the address of the cancer case is obtained via a spatial join from the estimated address-level groundwater radon estimates of Messier *et al.*³ The same confounding variables are included in the logistic cluster membership model as in the incidence rate model; however, differences due to the address-level information are present, which are explained in detail in the [supplementary material](#) (available as [Supplementary data at IJE online](#)). The odds ratio (OR), or the ratio of the odds that a case is a member of a cluster when a given predictor variable is increased by one unit, is calculated for each variable by exponentiating the logistic regression model coefficient. The OR does not directly reflect individual-level risk as in a classical case-control study design; however, individual-level explanatory variables provide additional evidence of associations with the cancer outcomes that is otherwise lost in an ecological study design. Similarly to the incidence rate model, we create and compare models with increasing levels of controlling for confounding variables. First, the crude model with only groundwater radon is developed. Second, the adjusted model controls for the effects of gender and age. And last, the full model controls for the additional factors of indoor air radon, smoking, race, public water supply, residential tenure, gender and age.

Spatial auto-correlation of model residuals is assessed by examining a spatial covariance plot of the model standardized Pearson residuals. If significant auto-correlation is present, which can potentially bias parameter and standard error estimates, then we implement a generalized estimating equation (GEE)^{31–34} which accounts for correlations between clusters and assumes no correlation within clusters (see [supplementary material](#), available as [Supplementary data at IJE online](#)). GLMs are modelled using the COUNT package²⁶ and GEEs are modelled using the GEE package³⁵ of the R statistical software. Spatial covariance of residuals are calculated using the *BMElib*²⁹ numerical toolbox in MATLAB. The cluster analysis was performed using the Cluster and Outlier Analysis tool in ArcGIS 10.0.³⁶

Results

The results for groundwater and indoor air radon in the crude, age- and gender-adjusted and full models of lung

cancer incidence rate, stomach cancer incidence rate, lung cancer cluster membership and stomach cancer cluster membership are summarized in [Table 2](#).

Lung cancer

The groundwater radon IRR and 95% confidence intervals are above 1 for all three lung cancer incidence rate models. The full model has an IRR of 1.03 (1.01, 1.06) indicating a 3% increase in lung cancer incidence rate for every 100 Bq/l increase in predicted census tract averaged groundwater radon concentration after controlling for confounding factors. Full results are available in the [supplementary material \(Table S2, available as Supplementary data at IJE online\)](#). Residual spatial-autocorrelation in the full lung cancer incidence rate model is considered insignificant based on the Pearson covariance plots ([supplementary Figure S3, available as Supplementary data at IJE online](#)).

The state-wide observed incidence for lung cancer during the study period is 95.7 and 52.8 cases per 100 000 person-years for males and females, respectively. This rate was used as the expected incidence in the cluster analysis of normalized excess cancer cases, which resulted in 254 out of 1554 (16.3%) census tracts having higher than expected rates of lung cancer ([Figure 1A](#)). A total of 13 414 (21%) cases occur within the clusters.

Address-level groundwater radon for the full lung cancer cluster membership GLM has an OR of lung cancer cluster membership of 1.32(1.29, 1.36; [Table S3, available as Supplementary data at IJE online](#)); however, it showed residual spatial auto-correlation. After implementing a GEE with an unstructured covariance matrix,³² spatial auto-correlation was reduced ([Supplementary Figure S3, available as Supplementary data at IJE online](#)) and the OR was reduced to 1.12 (1.04, 1.22; [Table 2](#)). The cluster membership GEE model indicates that predicted address-level groundwater radon is a risk factor for cluster

membership of lung cancer after controlling for confounding factors and spatial auto-correlation. Results for the all variables are in [Supplementary Table S3, available as Supplementary data at IJE online](#).

Stomach cancer

Groundwater radon IRR's for all three stomach cancer incidence rate models are above 1; however, the lower 95% confidence intervals are below 1. Full results are available in [Supplementary Table S4 \(available as Supplementary data at IJE online\)](#).

The state-wide observed incidence rate for stomach cancer during the study period is 8.2 and 4.1 cases per 100 000 person-years for males and females, respectively. This rate was used as the expected incidence in the cluster analysis of normalized excess cancer cases, which resulted in 113 out of 1554 (12.8%) census tracts having higher than expected rates of stomach cancer ([Figure 1B](#)). A total of 667 (12.8%) cases occur within the clusters.

Address-level groundwater radon for the full stomach cancer cluster membership GLM has an OR of stomach cancer cluster membership of 1.27 (1.14, 1.41; [Table S5, available as Supplementary data at IJE online](#)); however, it showed residual spatial auto-correlation. After implementing a GEE spatial auto-correlation was reduced ([Supplementary Figure S3](#)) and the OR was reduced to 1.24 (1.03, 1.49; [Table 2](#)). The cluster membership GEE model indicates that groundwater radon exposure is a risk factor for cluster membership of stomach cancer cases. Results for all variables are in [Supplementary Table S5](#).

Discussion

We presented ecological census tract incidence rates and case-only address-level cluster membership models for lung

Table 2. Summary of results for the NB2 incidence rate models, and the logistic cluster membership GEE models for lung and stomach cancers. For each model the crude, age- and gender-adjusted and full model IRR (95% confidence intervals) are presented for the intercept, groundwater radon and indoor air radon variables. Confounding variables are discussed in the text and available in full in the [supplementary material \(available at IJE online\)](#)

Model	Variable	Crude	Age- and gender-adjusted	Fully adjusted
Lung cancer incidence rate	Groundwater radon (per 100 Bq/l)	1.08 (1.03, 1.12)	1.03(1.01, 1.06)	1.03 (1.01, 1.06)
	Indoor air radon (per 100 Bq/m ³)			1.001 (0.99, 1.01)
Stomach cancer incidence rate	Groundwater radon (per 100 Bq/l)	1.06 (0.99, 1.13)	1.02(0.96, 1.08)	1.05 (0.99, 1.11)
	Indoor air radon (per 100 Bq/m ³)			0.98 (0.95, 1.004)
Lung cancer cluster membership (GEE model)	Groundwater radon (per 100 Bq/l)	1.13 (1.04, 1.23)	1.13 (1.04, 1.23)	1.12 (1.04, 1.22)
	Indoor air radon (per 100 Bq/m ³)			1.11(1.01, 1.21)
Stomach cancer cluster membership (GEE model)	Groundwater radon (per 100 Bq/l)	1.22 (1.02, 1.46)	1.21 (1.01, 1.45)	1.24 (1.03, 1.49)
	Indoor air radon (per 100 Bq/m ³)			1.09 (0.96, 1.23)

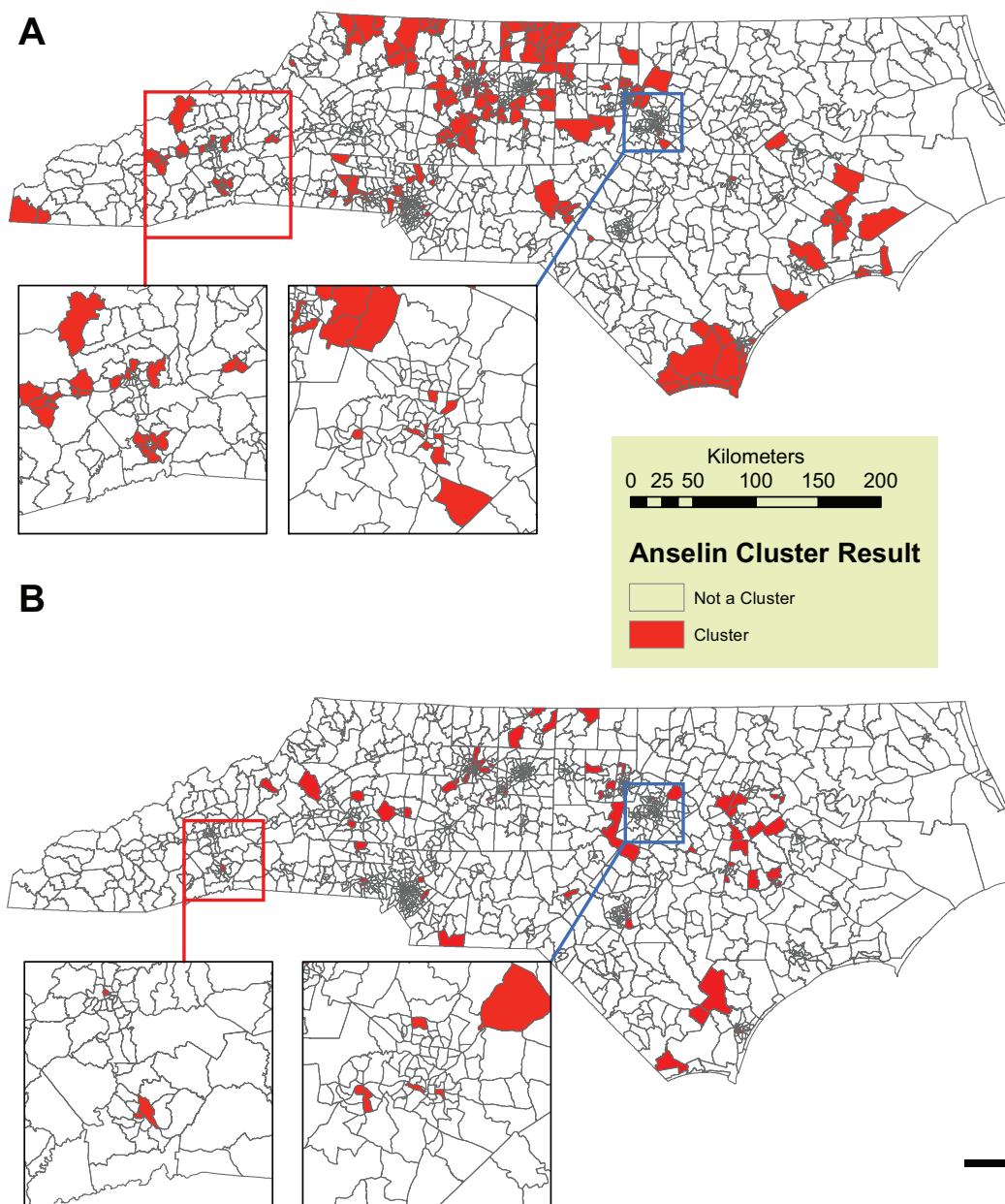


Figure 1. Anselin Local Moran's I clusters (Filled) for excess, normalized (A) lung cancer and (B) stomach cancer incidence calculated in ArcGIS 10.0. Cases are assigned a 1 status if they within a census tract identified a cluster. All other cases are assigned a 0 status. Each map has two inset maps of the Asheville (West Inset) and Raleigh (East Inset) metropolitan areas. Geomasked cancer cases are not shown, in order to protect spatial identity as per the data use agreement.

and stomach cancer in North Carolina, USA. Our goal was to quantify the associations between groundwater radon exposure and lung and stomach cancer, while considering not only the effects of known confounders, but also the spatial scale of outcome and explanatory variables. There have been several studies demonstrating that indoor air radon is associated with a significant risk for lung cancer,^{7,10-15} but there has been only one epidemiological study of groundwater radon exposure and lung cancer and this was an ecological study for mortality¹⁷ at the county

level that found positive and significant associations. There is general consensus on the biological and physical plausibility of groundwater radon leading to stomach cancer;^{8,9,16} however, there has only been one epidemiological study, with a small sample size and lack of control of confounders,⁴ to directly measure this association, which showed an insignificant association. Our study is the first epidemiological analysis finding a significant positive association between groundwater radon exposure and lung cancer incidence rates, and the first to find that an increase

of 100 Bq/l in groundwater radon concentration increases the odds that both lung cancer cases (OR = 1.12, 95% CI = 1.04, 1.22) and stomach cancer cases (OR = 1.24, 95% CI = 1.03, 1.49) are members of a cancer cluster, after controlling for confounding factors and spatial autocorrelation. The cluster membership model findings are made possible by the strength of our LUR-BME model quantifying exposure at the address level, which reduces the effects of the ecological fallacy.

Groundwater radon is a source of indoor air radon due to radon's transfer from water to air during showers,³⁷ laundry and washing dishes.⁸ The crude model result has an IRR of 1.08 (95% CI = 1.03, 1.12) per 100 Bq/m³; moreover, the age- and gender-adjusted model has an IRR of 1.03 (95% CI = 1.01, 1.06). We further investigate risks by controlling for confounding variables, which also results in an IRR of 1.03 (95% CI = 1.01, 1.06). Our incidence rate model results for lung cancer provide the first epidemiological evidence of increased lung cancer incidence from groundwater radon exposure alone. Compared with previous studies that consider groundwater radon as a contribution to indoor air radon,^{7,14} our study finds an increased risk from groundwater radon when considered separately. This is important because the methods for mitigating radon exposure from indoor air and groundwater may require different preventative measures.

The effect of groundwater radon concentration on lung cancer cluster membership was also positive with an OR of 1.12 (95% CI = 1.04, 1.22) for the full model; thus for every 100 Bq/L increase in groundwater radon concentration, after controlling for all confounding variables and spatial auto-correlation, there is a 12% increase in the odds that a lung cancer case occurs within a local lung cancer cluster. Since we have a case-only study design, the OR does not have the usual interpretation of an individual's odds of disease given an exposure compared with the odds of disease with no exposure; however, it does maintain an interpretation that reflects the underlying risk. In this two-stage analysis procedure, the clusters may represent regions with unknown underlying geographical risk factors for lung cancer, and the subsequent logistic regression analysis of case cluster membership suggests increased groundwater radon concentration may be one explanation of clustering, since it has an OR and 95% confidence interval greater than 1. It follows that our cluster membership GLM/GEE results supplement our census tract ecological study in providing the epidemiological evidence that groundwater radon concentrations result in an increased risk of lung cancer; and more importantly, the cluster membership model shows this based on a fine resolution model of exposure that captures the variability of address-level groundwater radon within each census tract, which is

important for radon since it is known to have significant local variability. Overall, our results for groundwater radon and lung cancer associations provide epidemiological evidence and support the National Research Council⁸ assessment of increased risk of lung cancer from groundwater radon exposure.

Lung cancer from indoor air radon exposure is the most well-studied target organ and pathway combination for radon.^{7,8,12-14,38-40} There is a general consensus that residential exposure from indoor air radon increases risk of lung cancer. This result was corroborated in the lung cancer cluster membership GEE model, which results in an OR of 1.11 (95% CI = 1.01, 1.21) for every 100 Bq/m³ of modelled census tract geometrical mean indoor air radon.

The effect of groundwater radon on stomach cancer is unclear in the incidence rate model results, with an IRR greater than 1 but lower 95% confidence bounds below 1 (Supplementary Table S4). Contrarily, the stomach cancer full cluster membership GEE model for groundwater radon has an OR with 95% confidence bounds above 1 for stomach cancer cluster membership. As previously mentioned, there is local variability in groundwater radon measurements that is likely diluted from census tract averaging, and subsequently makes finding a pronounced effect in the ecological incidence rate model more difficult. Additionally, the importance of accounting for residual spatial-autocorrelation is evidenced by the fact that there is a difference in groundwater radon OR between the adjusted logistic GLM and the adjusted logistic GEE for stomach cancer cluster membership (for groundwater radon: GLM OR = 1.27, GEE OR = 1.24; see Table S5).

The cluster membership GEE model shows that groundwater radon exposure is associated with increased risk for stomach cancer with a 24% increased odds that a stomach cancer case is member of a cancer cluster for every 100 Bq/l increase in concentration while controlling for all confounding factors and spatial auto-correlation. Our results provide epidemiological evidence that groundwater radon is an environmental risk factor for stomach cancer cases occurring within local stomach cancer clusters, which supports the National Research Council⁸ that groundwater radon is a significant risk for stomach cancer, but disputes the Auvinen *et al.* finding of no significant effects of radon exposure on stomach cancer.⁴ Auvinen *et al.* also find that uranium (the parent element to radon and a source of ionizing radiation) in drinking water has an insignificant but protective effect, which also contradicts the positive association Wilkinson *et al.*⁴¹ found between uranium deposits and stomach cancer incidence. Furthermore, Kjelberg and Wiseman¹⁸ found significant positive associations between indoor air radon and stomach cancer incidence.

The effects of other confounding variables are generally consistent with the literature, although there are some differences between the incidence rate and cluster membership models. For instance, we found males to have an IRR of 1.73 (95% CI=1.70, 1.77) and 1.94 (95% CI=1.83, 2.05) for lung and stomach cancer incidence rates, respectively, which is consistent with observed lung cancer⁴² and stomach cancer^{43,44} rates. Our results for gender provided no evidence of association at the 95% confidence level for either lung or stomach cluster membership models. Likewise, we found age 65 and over compared with age 64 and under to have an IRR of 13.5 (95% CI=13.2, 13.8) and 12.9 (95% CI=12.1, 13.7) for lung and stomach cancer incidence rates, respectively, which is also consistent with observed rates.^{42–44} We do not expect the gender and age variables to effect the groundwater radon results as they are uncorrelated with groundwater use and radon concentrations. We find the IRR for smoking prevalence and lung cancer to be 26.0 (95% CI=19.6, 34.6) which corresponds well to the results that smoking is the largest risk factor for increased lung cancer incidence.⁴² The results for smoking in the lung cancer cluster membership are inconclusive, which is likely due to the prevalence rates being assigned to the individual level. Moreover, we find the IRR of smoking and stomach cancer to be 3.2 (95% CI=1.6, 6.5), which has the same direction but larger magnitude compared with the rate ratio reported in Crew and Neugut⁴³ (2.1, 95% CI=1.2, 3.6) and the hazard ratio reported in Gonzalez *et al.*⁴⁵ (1.73, 95% CI=1.06, 2.83). Similar to the lung cancer cluster membership model, our stomach cancer cluster membership model has a result contrary to the literature; however, it is also likely because the census tract smoking prevalence does not adequately account for individual-level smoking. All of the full model variable results and interpretations are available in the [supplementary material](#) online.

Our results support the association between groundwater radon exposure and stomach cancer, which has been under-studied and has mixed results. Potential omitted confounders include *Helicobacter pylori* incidence, diet, family history, work in the rubber and fish hatchery industries and other radiation exposures. *Helicobacter pylori* is associated with untreated private well use^{46,47} but, since it occurs due to faecal contamination whereas radon is geological in nature, it is reasonable to assume that the presence of both radon and *H. pylori* in groundwater is uncorrelated. Likewise, the other omitted variables are likely uncorrelated with groundwater radon concentrations.

Limitations of the incidence rate models are normal for ecological studies, which include assigning exposures to an analysis unit area when it is known the exposure varies significantly at the individual level. The cluster membership

models improved upon this; however, there were still some controlling ecological-level variables assigned to individual cancer cases which may lead to residual confounding due to the differences in scale of the confounding and dependent variables. This residual confounding is mitigated by the cluster membership GEE model if it manifests in residual spatial-autocorrelation, which is likely since both residual confounding and spatial auto-correlation occur from non-fully specified models. Moreover, alternative models for identifying clusters are possible such as Geis-Ord G⁴⁸ or likelihood ratio scan statistics.⁴⁹ Nonetheless, our study should provide not only evidence of the associations but also the results needed to calculate the sufficient sample size needed to design a larger, individual-level epidemiological analysis such as a retrospective case-control or a prospective case-cohort study.

In summary, our study developed models for lung and stomach cancer associations with groundwater radon at the ecological scale with negative binomial regression and at the address level with logistic regression of case membership in cancer clusters. We find epidemiological evidence of the association between groundwater radon exposure and increased risk of lung cancer incidence while controlling for confounders at the ecological level. This is also the first epidemiological analysis to find groundwater radon to be a significant environmental risk factor underlying both lung and stomach cancer clusters.

Supplementary Data

[Supplementary data](#) are available at IJE online.

Funding

This research was supported in part by funds from the NIH T32ES007018, NIOSH 2T42OH008673.

Acknowledgements

We thank the North Carolina Central Cancer Registry and the State Center for Health Statistics, including Dr Chandrika Rao, Dr Luis Carrasco and Christian Klaus for providing, geocoding and geomasking the cancer data and for their comments on the manuscript.

Conflict of interest: None declared.

References

1. Loomis DP. Radon-222 concentration and aquifer lithology in North Carolina. *Groundw Monit Remediat* 1987; (Spring):33–39.
2. Yang Q, Smitherman PE, Hess CT *et al.* Uranium and radon in private bedrock well water in Maine: geospatial analysis at two scales. *Environ Sci Technol* 2014;**48**:4298–306.

3. Messier KP, Campbell T, Bradley P, Serre ML. Estimation of groundwater radon in North Carolina using Land Use Regression and Bayesian Maximum Entropy. *Environ Sci Technol* 2015;49:9817–25.
4. Auvinen A, Salonen L, Pekkanen J, Pukkala E, Ilus T, Kurttio P. Radon and other natural radionuclides in drinking water and risk of stomach cancer: a case-cohort study in Finland. *Int J Cancer* 2005;114:109–13.
5. Zhu HC, Charlet JM, Poffijn A. Radon risk mapping in southern Belgium: an application of geostatistical and GIS techniques. *Sci Total Environ* 2001;272:203–10.
6. De Francesco S, Tommasone FP, Cuoco E, Verrengia G, Tedesco D. Radon hazard in shallow groundwaters: amplification and long term variability induced by rainfall. *Sci Total Environ* 2010;408:779–89.
7. Darby S, Hill D, Auvinen A *et al.* Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. *BMJ* 2004;330:223.
8. National Research Council. *Risk Assessment of Radon in Drinking Water*. Washington, DC: National Academies Press, 1999.
9. Campbell T, Mort S, Fong F *et al.* *North Carolina Radon-in-Water Advisory Committee Report*. Raleigh, NC: North Carolina Radon-in-Water Advisory Committee, 2011.
10. Field RW, Smith B, Steck D, Lynch CF. Residential radon exposure and lung cancer: variation in risk estimates using alternative exposure scenarios. *J Expo Anal Environ Epidemiol* 2002;12:197–203.
11. Kendall GM, Smith TJ. Doses to organs and tissues from radon and its decay products. *J Radiol Prot* 2002;22:389–406.
12. Lubin JH, Boice JD. Lung cancer risk from residential radon: meta-analysis of eight epidemiologic studies. *J Natl Cancer Inst* 1997;89:49–57.
13. Field RW. Environmental factors in cancer: radon. *Rev Environ Health* 2010;25:33–38.
14. Krewski D, Lubin JH, Zielinski JM *et al.* Residential radon and risk of lung cancer. *Epidemiology*. 2005 Mar;16:137–45.
15. International Agency for Research on Cancer. *Monographs on the Evaluation of Carcinogenic Risks to Humans: Radiation*. Geneva: World Health Organization, 2012.
16. Crawford-Brown DJ. Cancer fatalities from waterborne radon (Rn-222). *Risk Anal* 1991;11:135–43.
17. Hess CT, Weiffenbach CV, Norton SA. Environmental radon and cancer correlations in Maine. *Health Phys* 1983;45:339–48.
18. Kjelberg S, Wiseman JS. The relationship of radon to gastrointestinal malignancies. *Am Surg* 1995;61:822–25.
19. Centers for Disease Control and Prevention. *U.S. Cancer Statistics: An Interactive Atlas*. 2015. http://apps.nccd.cdc.gov/DCPC_INCA/DCPC_INCA.aspx. (3 January 2015, date last accessed).
20. Gilbert M, Xiao X, Pfeiffer DU *et al.* Mapping H5N1 highly pathogenic avian influenza risk in Southeast Asia. *Proc Natl Acad Sci U S A* 2008;105:4769–74.
21. Loth L, Gilbert M, Osmani MG, Kalam AM, Xiao X. Risk factors and clusters of highly pathogenic avian influenza H5N1 outbreaks in Bangladesh. *Prev Vet Med* 2010;96:104–13.
22. Torbick N, Hession S, Stommel E, Caller T. Mapping amyotrophic lateral sclerosis lake risk factors across northern New England. *Int J Health Geogr* 2014;13:1.
23. Tsai P-J, Lin M-L, Chu C-M, Perng C-H. Spatial autocorrelation analysis of health care hotspots in Taiwan in 2006. *BMC Public Health* 2009;9:464.
24. Allshouse WB, Fitch MK, Hampton KH *et al.* Geomasking sensitive health data and privacy protection: an evaluation using an E911 database. *Geocarto Int* 2010;25:443–52.
25. Hampton KH, Fitch MK, Allshouse WB *et al.* Mapping health data: improved privacy protection with donut method geomasking. *Am J Epidemiol* 2010;172:1062–69.
26. Hilbe JM. *Negative Binomial Regression*. 2nd edn. Cambridge, UK: Cambridge University Press, 2011.
27. Messier KP, Jackson LE, White JL, Hilborn ED. Landscape risk factors for Lyme disease in the eastern broadleaf forest province of the Hudson River valley and the effect of explanatory data classification resolution. *Spat Spatiotemporal Epidemiol* 2015;12:9–17.
28. Ortega Hinojosa AM, Davies MM, Jarjour S *et al.* Developing small-area predictions for smoking and obesity prevalence in the United States for use in environmental public health tracking. *Environ Res* 2014;134:435–52.
29. Christakos G, Bogaert P, Serre ML. *Temporal GIS: Advanced Function for Field-Based Applications*. New York, NY: Springer, 2002.
30. Sanders AP, Messier KP, Shehee M, Rudo K, Serre ML, Fry RC. Arsenic in North Carolina: public health implications. *Environ Int* 2011;38:10–16.
31. Liang K-Y, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;73:13–22.
32. Carl G, Kühn I. Analyzing spatial autocorrelation in species distributions using Gaussian and logit models. *Ecol Modell* 2007;207:159–70.
33. Dormann CF, McPherson JM, Araújo MB *et al.* Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography (Cop)* 2007;30:609–28.
34. Dormann CF. Assessing the validity of autologistic regression. *Ecol Modell* 2007;207:234–42.
35. Carey VJ, Lumley T, Ripley B. *Generalized Estimation Equation Solver*. 2012. https://cran.r-project.org/web/packages/Ch_analysing_longitudinal_dataII.pdf
36. Environmental Systems Research Institute. ArcGIS. Redlands, CA; 2012.
37. Vinson DS, Campbell TR, Vengosh A. Radon transfer from groundwater used in showers to indoor air. *Appl Geochemistry* 2008;23:2676–85.
38. Pavia M, Bianco A, Pileggi C, Angelillo IF. Meta-analysis of residential exposure to radon gas and lung cancer. *Bull World Health Organ* 2003;81:732–38.
39. Darby SC, Whitley E, Howe GR *et al.* Radon and cancers other than lung cancer in underground miners: a collaborative analysis of 11 studies. *J Natl Cancer Inst* 1995;87:378–84.
40. Field RW, Withers BL. Occupational and environmental causes of lung cancer. *Clin Chest Med* 2012;33:681–703.
41. Wilkinson GS. Gastric cancer in New Mexico counties with significant deposits of uranium. *Arch Environ Health* 1985;40:307–12.
42. Alberg AJ, Samet JM. Epidemiology of lung cancer. *Chest* 2003;123:21S–49S.

43. Crew KD, Neugut AI. Epidemiology of gastric cancer. *World J Gastroenterol* 2006;**12**:354–62.
44. Brenner H, Rothenbacher D, Arndt V. Epidemiology of stomach cancer. *Methods Mol Biol* 2009;**472**:467–77.
45. Gonzalez CA, Pera G, Agudo A *et al.* Smoking and the risk of gastric cancer in the European Prospective Investigation Into Cancer and Nutrition (EPIC). *Int J Cancer* 2003;**107**:629–34.
46. Brown LM. Helicobacter pylori: epidemiology and routes of transmission. *Epidemiol Rev* 2000;**22**:283–97.
47. Baker KH, Hegarty JP. Presence of Helicobacter pylori in drinking water is associated with clinical infection. *Scand J Infect Dis* 2001;**33**:744–46.
48. Getis A, Ord JK. The analysis of spatial association by use of distance statistics. *Geogr Anal* 1992;**24**:189–206.
49. Kulldorff M. A spatial scan statistic. *Commun Stat Theory Methods* 1997;**26**:1481–96.