Modeling Joint Exposures and Health Outcomes for Cumulative Risk Assessment: The Case of Radon and Smoking

Teresa Chahine [1,*], Bradley D. Schultz [2], Valerie G. Zartarian [2], Jianping Xue [2], SV Subramanian [1] and Jonathan I. Levy [1,3]

[1] Harvard School of Public Health, Harvard University, 677 Huntington Avenue, Boston, MA 02215, USA; E-Mails: svsubram@hsph.harvard.edu (SV.S.); jonlevy@bu.edu (J.I.L.)

[2] US Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory, Research Triangle Park, NC 27711, USA; E-Mails: schultz.brad@epa.gov (B.D.S.); zartarian.valerie@epa.gov (V.G.Z.); xue.jianping@epa.gov (J.P.X.)

[3] Boston University School of Public Health, 715 Albany Street, Talbot Building, Boston, MA 02118, USA

ABSTRACT (Word count = 230)

Residential radon exposure is estimated to cause 20,000 lung cancer deaths annually, but concentrations are highly variable and communities assessing their environmental health hazards are in need of more geographically refined estimates Measurement of radon in each home is recommended by the U.S. Environmental Proection Agency (EPA) but communities assessing local risks often overlook radon without an idea of how its risks compare to other risks in their communities . Previous studies have estimated radon concentrations by county, but no study has estimated county-level radon human exposures and risks across the country. This study assesses whether publicly-available sociodemographic and geographic variables from the 2000 U.S. Census can characterize variance from the national average risk estimate provided by the EPA, by providing multilevel models employing sociodemographic and geographic predictors to estimate variability in radon concentrations and in smoking, an effect modifier. A strength conferred by multilevel modeling is that spatial correlations in radon levels and smoking prevalence are accounted for. Using the estimates of estimated fatal lung cancer risk from residential radon exposure by county, we find that even in counties with low estimated radon concentrations, the estimated average population risk still exceeds 1 in 1000. . While the predictive power is modest given limited national-level data about local soil factors and other covariates affecting radon, this risk model considers radon in the context of cumulative lung cancer risk and leverages available data to improve screening-level community-scale risk characterizations.

(Word count = 6,094)

1. INTRODUCTION

Radon consistently emerges as a significant concern for public health and a high environmental health priority when viewed from a human risk perspective. Radon is the leading cause of lung cancer deaths in non-smokers, and the second leading cause of lung cancer deaths in smokers[1]. The United States Environmental Protection Agency (EPA) has estimated approximately 20,000 lung cancer cases annually, with an average lifetime risk of fatal lung cancer of 0.73% in the U.S. general population based on the national average concentration of 1.25 pCi/L.[1] While the nationally-aggregated estimates of health risks from radon has been informative, communities and individuals concerned with radon have expressed the need for more refined spatial information on radon concentrations, exposure, and risk. Some community members view their communities as "different" than the typical community and sometimes discount national estimates. Additionally, some underestimate the impact of radon health effects, when their area is deemed "average" or "below average" in likely radon concentration; on a risk basis, even lower risk areas can have risks exceeding those associated with issues of higher current visibility.

There is an increasing call for addressing risks cumulatively, that is, to include the multiple stressors related to the health impact[2,3]. The main factors affecting lung cancer risk among the general population are residential radon concentrations and smoking: current and former smokers have a higher unit risk for lung cancer per unit of radon exposure than non-smokers[4,5]. Both residential radon concentrations and smoking prevalence are highly variable across different locations and different populations in the U.S. Previous studies have examined each separately, but no study has jointly evaluated the demographic and geographic patterning of variables associated with radon and smoking and the subsequent patterning of radon risk.

The distribution of residential radon concentrations across the U.S. has been the subject of numerous studies for the past several decades, and is related to both geological and housing characteristics. Radon originates from radium in underlying bedrock, the composition of which is determined by rock type and origin. Radon travels through soil and infiltrates built structures through cracks, cavities and construction joints[6]. Soil type, texture, moisture and permeability affect the movement of radon gas, in combination with climate and meteorology[7,8]. One of the primary drivers of the movement of radon from the soil into the indoor environment is pressure gradients, which can be caused by temperature differences, wind, and building heating or ventilation[9,10,11].

Due to the complex interplay of the factors described above and the lack of data on soil permeability to gas, it can be challenging to model radon concentrations, and previous investigations have had some limitations. The U.S. Geological Survey assigned a radon potential score by geological province based on expert evaluation of available geological and soil surveys, but could not capture local variability in soil and housing factors due to lack of local data on these factors[12]. The EPA added to the above score by incorporating measured concentrations and architecture information from state residential radon surveys to produce a national zone map of estimated radon levels by county[13]. However, state databases are of varying quality and present considerable challenges for developing nationally-consistent radon concentration estimates; in addition, they are largely based on short-term screening measurements which are limited for providing the long-term estimates needed to determine lung cancer risk. Long-term measurements are available in a limited number of state surveys, but are most well represented nationally in the National Residential Radon Survey (NRRS)[14]. A study using measurements from the NRRS estimated median long-term residential radon concentrations by county, but cautioned that variability within a county could be significant and did not include information necessary to link with smoking data[15]. More generally, because of the high level of variability and the lack of sufficient data on factors affecting radon concentrations at fine spatial resolution, numerous studies such as the above have reported limited predictive power in modeling radon concentrations, posing challenges for more geographically and demographically refined assessments.[16,17,18,19,20,21] While no model can eliminate the need for radon measurements in each home, especially given the high risks from radon, models have promise for providing screening-level estimates for community groups and individuals in understanding the relative importance of radon in their communities.

For smoking, there are similar challenges in comprehensively incorporating geographic and demographic variability. Smoking statistics are available at a national and state level, and local data are available in some communities but are not systematically collected and reported across the U.S. Variability in smoking is related to compositional factors (individual demographic characteristics and socioeconomic indicators) and contextual factors (neighborhood characteristics, local and state legislations)[22,23,24,25]. Previous studies have quantified the association between smoking and compositional and contextual factors across different populations and places, but no studies have provided models with sufficient geographic and demographic stratification and coverage to allow for a refined examination of lung cancer risks from residential radon exposure in the U.S.

Despite the challenges of a cumulative assessment of radon and smoking, community groups have repeatedly asked for cumulative assessments, and the smoking and radon cumulative effect is one of the best understood and tractable interactions[Error! Bookmark not defined.,26]. In this study, we develop a systematic approach to model both radon and smoking at high spatial and demographic resolution across the U.S., linking multiple national databases and accounting for spatial correlations using a multilevel modeling framework[27]. We construct a multilevel regression model predicting radon concentrations using only sociodemographic and geographic covariates that can be included in a multilevel regression model predicting smoking, in order to link the two in a community-scale risk assessment. We use predictors which are available across the U.S. from the Census, and leverage components of the EPA national risk assessment to develop a framework to provide communities and decision-makers with more spatially refined estimates of lung cancer risk from residential radon exposure.

2. METHODS

2. 1. Conceptual Framework

To capture variability in parameters affecting radon risk, we built three statistical regression models using three national datasets described below. The first model provides estimates of residential radon concentrations based on locational information and house type; the second model provides estimates of house type based on occupants' sociodemographic characteristics, and the third model provides estimates of individual smoking status based on the same sociodemographic characteristics. Together, these regression models provide estimates for the association of sociodemographic and geographic variables with radon concentration, housing characteristics, and smoking prevalence (Table I). Resulting parameter estimates from the three models can be combined to produce radon exposure estimates and smoking prevalence estimates for a location (e.g., county or census tract) based on its geography and composition.

Our objective is to provide estimates of risk across multiple sociodemographic and geographic subpopulation bins. The U.S. Census provides cross-tabulated data on the number of people by age, sex, race, and poverty status, at levels of geographic resolution down to the census tract (small statistical subdivisions of a county, usually containing between 2500-8000 persons)[28]. Census tracts are therefore relatively small geographic entities with sufficient population size to yield cross-tabulated demographics, and were also designed to be homogeneous with respect to population characteristics, economic status, and living conditions. Within the present analysis, we present all exposure and risk calculations at county resolution for ease of presentation and proof of concept, but based our analysis plan on providing smoking and radon concentration estimates at the census tract level using these cross-tabulated variables in addition to geographic variables as applicable in each model.

2.2. Radon Concentration Model

We developed our radon model from the NRRS, during which long-term measurements of radon concentration were taken in all living levels of a nationally representative sample of homes from 1989-1990[11]. While somewhat outdated with respect to current housing stock, it represents the most robust and geographically representative data set publicly available. Information on housing characteristics collected in the NRRS was combined with data from the USGS and soil surveys based on the location of each home, which was then discarded for confidentiality reasons[29]. We developed a log-linear model to quantify associations at the household level between geological and meteorological variables, housing characteristics, and annual average radon concentrations averaged over all living levels. Radon concentrations were scaled to adjust negative measurements recorded by the survey instruments to match minimal outdoor radon concentrations based on methods previously published by Price et al[12]. Analysis was conducted using MLwiN 2.16[30].

To account for the geographic clustering of samples in the NRRS we built a four-level model of housing units (n=5336) nested within secondary sampling units (SSU, n=977) nested within primary sampling units (PSU, n=125) nested within states (n=44). Indicator variables were used to represent Census Region. In the survey primary sampling units corresponded to one or more counties, and secondary sampling units corresponded to census tracts or parts of census tracts within the sampled counties.

We selected potential covariates based on results of previous statistical analysis of the NRRS data[31]. At the county level, we included meteorological variables from a national meteorological database[32]. Alaska was excluded from the analysis because meteorological data were not available for this state within the database most appropriate for this assessment. Geological variables (soil texture, permeability, underlying bedrock, equivalent uranium) were evaluated both separately and using a summary score of geological radon potential provided by the

USGS[33]. The USGS score was provided within geological province boundaries; we assigned a score to each county based on the geological province in which the county is located, and for those counties located at the intersection of more than one geological province we assigned the score of the province which covered the largest area of the county.

At the household level, while numerous home characteristics would theoretically be linked with residential radon levels, our model structure necessitated that we restrict potential covariates to those available in the American Housing Survey (AHS)[34]. The AHS is the primary source of data for the U.S. housing stock. We tested for statistically significant associations between housing variables and log radon concentrations in univariate and multivariable models, and performed chi-square tests to measure correlations between the housing variables. We then assessed the predictive power of different housing variables by comparing the reduction in variance at the state, PSU, and SSU levels in different models, as well as overall fit using log-likelihood ratio tests.

2.3. Housing Model

In order to apply the radon concentration model across the U.S., we needed to link housing characteristics predictive of residential radon with sociodemographic and geographic data available in all locations. We developed a multinomial logistic regression model to quantify associations between housing type and publicly available sociodemographic and geographic covariates from the U.S. Census. As a result of the model-building described above, housing type was divided into five mutually exclusive categories: attached unit or mobile home, single detached unit with basement, single detached unit with crawl-space, single detached unit with slab-on-grade, and other single detached unit. Analysis was conducted using SAS 9.2 (*SAS* Institute Inc.). Because the dependent variable has the same value for all individuals nested within a household, a multilevel model is not possible, thus the clustering of individuals within households is not accounted for in our model. State and county identifiers were not provided in the AHS dataset; metropolitan statistical area was identified for less than half of the houses and was thus not included in our analysis. Therefore Census region was the only geographic covariate.

2.4. Smoking Model

We used a multilevel logistic modeling approach to develop predictors of smoking, using data from the 2006-7 Current Population Survey – Tobacco Use Supplement (CPS-TUS). This approach has been described in detail elsewhere[35]. For the purposes of the current study, the binomial outcome modeled was ever-smoking rather than current smoking only, as the unit risk for lung cancer from radon exposure differs for non-smokers compared to ever-smokers. Covariates were: individual-level variables available from census cross-tabulations (age, sex, poverty, race), area poverty at the CBSA (core-based statistical area) level, and tax laws and legislation at the state level. Analysis was conducted in MLwiN 2.16.

2.5. Exposure and Risk Estimates

Census 2000 Summary File 3 tables were obtained to provide the number of people in each sociodemographic bin (as defined by age, sex, race, and poverty status) in each county in the U.S. Because our smoking model was based on an adult study population, we included only individuals aged 18 and above in the risk calculations.

To estimate radon risk, we first determined the predicted probability of ever-smoking for each sociodemographic bin in each county by summing fixed effects of age, sex, race, poverty status, CBSA poverty, state tax, state legislation, and previous state smoking prevalence, in addition to state and CBSA residuals. Subpopulations with Black race were also assigned state-specific effect estimates for race. 180 CBSAs were not included in the CPS sample, and only state residuals were applied for these. Second, the predicted probability of each housing type was calculated for each bin based on the housing model by summing fixed effects of age, sex, race, poverty status, and Census region. Third, for each housing type, the predicted radon concentration was calculated for each bin by summing fixed effects of age, sex, race, poverty status, county meteorological variables, radon geological potential score, and state residuals (as well as county residuals for the 125 counties that were included in the NRRS sample). Five states were not included in the NRRS sample and were assigned zero residuals. Thus, we obtained estimates of the prevalence (predicted probability) of ever-smokers and of the five different housing types and corresponding radon concentrations for each sociodemographic bin in each county.

Based on this information, we estimated the population average risk for each county, following the EPA risk assessment algorithm and assuming that there is no differential distribution of non-smokers and current/former smokers among the different housing types within each geographic and demographic subpopulation. We applied an exposure rate of 0.144 working level months (WLM) per year for each pCi/L of radon gas, assuming that on average people spend about 70% of their time indoors at home, and that the equilibrium fraction for radon progeny is 40%[36].

WLM is the cumulative exposure measure used in the epidemiologic literature on uranium miners, from which the unit risk factors for lung cancer were derived[2]. The unit risk factors per WLM are 0.00106 and 0.000851 for male and female ever-smokers; 0.000174 and 0.000161 for male and female nonsmokers, respectively[1].

3. RESULTS

3.1. Radon Concentration Model

At the county level, the USGS summary score had higher statistical significance and improved the fit of the model more than the separate geological and soil covariates (as assessed using log-likelihood ratio tests). Annual heating infiltration degree days and average diurnal temperature difference were retained as meteorological variables (Table II). At the house level, statistically significant variables which improved the fit of the model and were available in the AHS dataset were: type of unit (detached vs. attached), presence of basement, presence of central air conditioning, use of gas fuel for heating, use of steam or hot water distribution system for heating, number of gas appliances, and year built. However, chi-square tests showed multiple correlations between these housing variables, and the use of numerous housing variables complicates linkages with individual census data. We fit a model containing a five-category house type variable (type of unit, basement) and it explained 85% of between-state variance, 51% of between-county variance, and 25% of between-census tract variance, compared to a model including all housing variables which explained 86% of between-state variance, 50% of between-county variance, and 29% of between-census tract variance. We therefore utilized the five-category house type variable in subsequent analyses.

Census region was a statistically significant predictor (p = 0.005), although with no statistically significant differences among the South, West, and Midwest. All county-level and house-level covariates in the final model were significant, with the exception of the crawl-space indicator in the house type variable.

3.2. Housing Model

House types were significantly associated with Census region, poverty status, age, and race. (No significant differences were observed by gender.) White subpopulations living above the poverty threshold in the Midwest had the highest odds of living in detached homes with basements compared to attached homes. Subpopulations with the lowest odds of living in detached homes with basements compared to attached homes were Black race, below poverty threshold, ages 25-34, living in the West. Parameter estimates for the housing model are presented in Appendix I.

3.3. Smoking Model

The prevalence of ever-smokers in the CPS-TUS 06-07 was 38.6% (17.9% current smokers, 20.7% former smokers). Associations of sociodemographic variables with ever-smoking were comparable to the associations reported previously for current smoking prevalence, with a few exceptions. The inverted U-shaped association for age peaked at a higher age than in the model for current smoking prevalence. State legislation restricting smoking in public venues and percent poverty at the CBSA level were not significant predictors of ever-smoking and did not show the same directionality as for the previously published current-smoking model. State cigarette excise tax showed a significant negative association with ever-smoking; this association persisted after controlling for previous state smoking prevalence, and is therefore not likely due to endogeneity or reverse causation. Men showed higher odds of smoking than women, and this effect was modified by race. The variance of the random parameters at the state and CBSA levels were 0.005 and 0.040 respectively, compared to 0.004 and 0.013 in the previously published current smoking model with unconstrained covariates. Parameter estimates for the smoking model are listed in Appendix II. These results are consistent with published results obtained from the CDC website for CPS-TUS data[37].

3.4. Exposure and Risk Estimates

County average radon lung cancer risk estimates ranged from 0.15% to 1.8%, with a mean by county of 0.66% and a median of 0.64% (standard deviation = 0.3%). High-risk clusters were observed in the northern Midwest states, which had relatively high levels of both radon and ever-smoking (Figure 1). South Dakota in particular shows a number of counties which contained among the highest estimated mean radon concentrations, and the same counties were also on the higher end of estimated ever-smoking prevalence. Two of the six counties nationwide which show predicted mean concentrations greater than 4 pCi/L were observed in Utah; however, because Utah has among the lowest smoking rates in the country, these counties did not emerge among the highest risk counties in the risk map. High-smoking clusters were observed in selected states in the Midwest and Southeast, where radon concentrations were on the lower end, and therefore risk clusters did not emerge in these states. Missouri and Kentucky in particular were among the states with the highest predicted probability of ever-smokers, but had average radon risk levels. Coastal states had the lowest radon concentrations, and many of these were also below-average smoking states, therefore resulting in the lowest risk.

4. DISCUSSION

The interplay between radon concentrations and smoking, and the demographically and geographically variable nature of both, results in a spatial distribution of risks that has not been captured in previous studies. Although there is still considerable uncertainty in the estimates, the models are capturing a wide range of risks associated with the highly variable predictors. The approach here allows for a combined radon and smoking lung cancer prediction. For example, jointly examining the patterns of demographic and contextual predictors associated with radon and smoking would allow for identifying the location of clusters with the highest predicted probability of fatal lung cancer from residential radon exposure.

Comparing the risk map and concentration map shown in Figure 1, the patterns follow a similar trend in many places but are far from identical. The shifting of patterns between the two maps can be illustrated by comparing areas with similar radon concentrations but different smoking patterns; for example, while Indiana does not stand out as a high radon area in the concentration map relative to Utah, its risk levels are higher due to the large difference in smoking prevalence between the two states. Comparing our risk map with previous screening maps such as the EPA map of radon zones, the overall patterns agree but nuances emerge within the highest potential zone, as illustrated by the high risk cluster in the northern Midwest.

The underlying Census sociodemographic data behind our risk map form a key factor which contributes directly to the smoking predictions through the strong association between compositional variables and smoking prevalence, and indirectly to the concentration predictions through the housing model component. For example, white male populations in the Midwest region have higher odds of living in detached single units with basements and have higher odds of smoking. When such populations are located in counties with high geological radon potential and higher than average diurnal temperature swings and total infiltration heating degree days, their lung cancer risk from residential radon exposure will likely exceed the national average. On the other hand, multi-directionality in exposure and risk factors was also observed; for example, living below the poverty threshold was negatively associated with the presence of a basement, thus likely to have lower radon concentrations after controlling for location, while it was positively associated with probability of smoking. While the sociodemographic data employed in our model captures variability in housing types and smoking patterns with respect to the national average, many factors remain which influence these outcomes and which were not controlled for in our models. These include other socioeconomic indicators (for example education, occupation, immigrant status, marital status) as well as contextual factors (for example local variations in construction patterns and smoking restrictions). Additionally, it is important to note that our risk model assumes a stationary population. It is unknown how radon-related risks might be affected by the migration of individuals and populations from one part of the country to another.

Our models have a number of limitations and should be considered for screening purposes only. As in any statistical regression, statistical inferences are drawn about the true population distribution based on a limited number of samples. The NRRS is nationally representative but included samples from only 125 counties and does not capture housing structures built within the past 20 years. Although the radon concentration model benefited from a multilevel structure in which state and county effects were drawn from a random distribution, random parameter variance was not fully captured by the available data. Thus there remain unexplained state and county effects. The housing model did not benefit from a multilevel structure because of the nature of the outcome variable, which is the same for all individuals in a household. Among the predictors of the housing model, poverty status is the most relevant for interpretation because it is a shared household characteristic (although measured as individual level variables in this dataset). On the other hand, sex and age and race are individual characteristics, and the nesting of individuals within households was not accounted for due to the lack of multilevel structure. Another weakness of the housing model was lack of higher levels such as county, metropolitan statistical area, or state, which were not included due to limited geographic identifiers in the AHS.

Although radon is difficult to predict due to the local variation in soil factors for which no data are available, and our regression models used a more limited set of covariates than in previous investigations, our radon concentration predictions compare well with previous estimates. In addition, our model points to the predictive power of the USGS radon potential score compared to individual soil and rock variables collected in geological databases and soil surveys. This is in agreement with the complex relationships of the multiple geological parameters affecting radon concentrations, which may best be summarized in the expert evaluation process used by the USGS to assign radon potential zones to each geological province. We also quantified the contribution of key housing factors which were significantly associated with census cross-tabulated variables. Although many housing factors affect indoor radon concentrations as previously described, the multinomial house type variable retained in our final model performed well when compared to multiple individual and often highly correlated housing variables. The covariates in our ever-smoking model explained a comparable amount of state and CBSA variance to previously published smoking models, despite being constrained to publicly available covariates only. Finally, it is important to note that the unit risk coefficient used in our risk estimates was developed by the EPA based on 1990 mortality rates and smoking prevalence data; however, our smoking estimates were based on demographic data from 2000.

These results are not intended to replace individual home measurements, which are recommended by the EPA for almost all homes. Further developed and evaluated, however, this type of modeling could provide insight to communities seeking screening-level information on their radon exposure and risk. Even in the lowest risk counties, average lung cancer risk from residential radon exposure exceeds one in a thousand. Similar statistical models have been developed for individual use, but require data inputs on individual homes[38]. Our model leverages publicly available data to provide average risk estimates for sociodemographic subpopulations within counties or census tracts across the U.S. We presented county average estimates within this paper for ease of presentation, but the methodology would be computationally identical at the census tract level, with improvements in the predictive power and potential interpretability of our models (including for housing, where urban/rural status could be incorporated as a covariate). Given that county average estimates are a relatively crude measure for community risk characterization, we encourage the development of census tract level estimates using the models we have developed.

Future research can improve the strength of these models by considering variability in other parameters affecting exposure and risk, such as time spent at home (which may also vary by age, sex, poverty, and race) and equilibrium fraction (which depends on factors such as particle size distributions in homes, in turn affected by smoking patterns). Model evaluation is also needed using measured long-term residential radon concentrations within selected geographic areas and demographic subpopulations, or through evaluation of lung cancer risk patterns among non-smokers. Future work to characterize and quantify the uncertainty in the risk calculations would be warranted and could provide valuable information to end-users of these predictions. Model predictions can also be refined and pilot tested using available local field data including but not limited to radon measurements or community surveys on household characteristics. Finally, we recommend that national survey bureaus make efforts to provide increased geographic identifiers in public use data files such as the AHS, subject to confidentiality constraints, given the importance of location in determining variability in population exposure and risks.

Despite the inherent data gaps and limitations, our models provide an approach for leveraging publicly available information from nationally representative data to capture correlations among parameters affecting both residential radon exposure and subsequent risk of lung cancer. Our model yields similar national-scale radon lung cancer risk estimates as prior analyses (0.7%) but illustrates an order of magnitude variation in average risk at county resolution. This research is the first to quantify variability in lung cancer risk from residential radon exposure, moving beyond the national average estimates provided in the EPA national risk assessment. Our study can provide a key input to community-scale risk prioritization efforts.

Acknowledgements:

This research was supported by EPA Student Services Contract #EP09D000560. The authors would like to thank R. Randall Schumann and Phillip N. Price for sharing their previous work. We also gratefully acknowledge Dr. Andrew Geller of EPA's Office of Research and Development, National Exposure Research Laboratory, for providing technical review of the paper.

Table I.

Regression models used to capture variability in exposure & risk parameters:

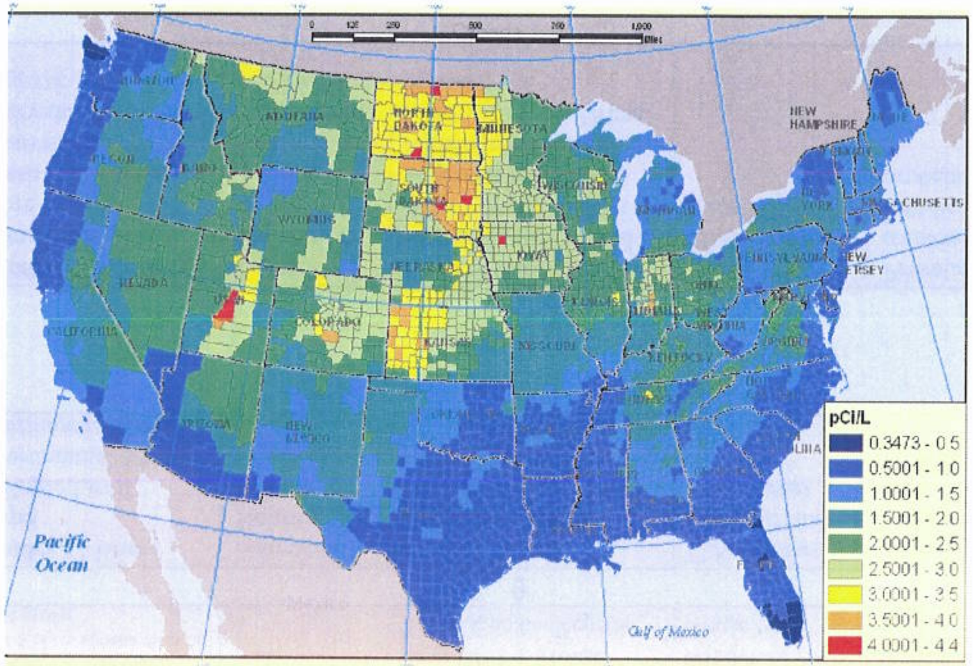
Regression Model (Multilevel Structure) ***and distributional assumptions***	**Dataset (n)**	**Covariates (Level)**	**Public Data Sources for Covariates**	**Predicted Outcome (units)**
Model #1 Radon Concentration (State, PSU*, SSU, House) *Log-Linear* *where PSU is county and SSU is census tract or bloc group	National Residential Radon Survey (5,413)	Radon potential category (PSU), Annual average diurnal swing (PSU), Annual total heating degree-days (PSU), House Type, Census Region	USGS radon potential score, LBNL meteorological database, Model #2 for house type	Annual average radon conc, averaged over all living levels, (log pCi/L), scaled
		⇧		
Model #2 House Type (Single Level Multinomial Model) *Multinomial Logistic*	American Housing Survey (39,107)	Poverty status, Age, Race, Sex, Census Region, Urban Status	U.S. Census 2000 Census-tract Tables	House type, 5 categories (predicted probability)
Model #3 Smoking Prevalence (State, Core Based Statistical Area, Household, Individual) *Binomial Logistic*	Current Population Survey – Tobacco Use Supplement (227,428)	Poverty status, Age, Race, Sex (individual), Area Poverty (CBSA), Cigarette taxes and legislations (State), Census Region, Urban Status	U.S. Census 2000 Census-tract Tables	Prevalence of smoking (predicted probability)

*Source: USEPA (2003) EPA assessment of risk from radon in homes. [EPA 402-R-03-003]

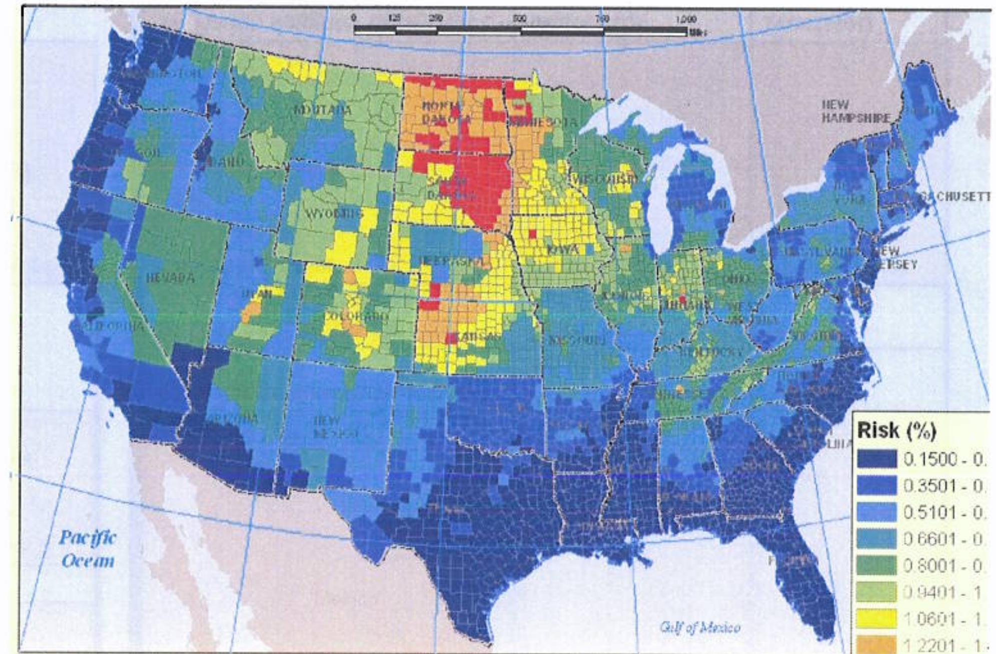
Risk Assessment Algorithm*: R = C x L x β x D x U x Y

Parameter	**Data Source**	**Geographic Resolution**
R: risk of fatal lung cancer from lifetime residential radon exposure	Product of below parameters	As per parameters below
C: radon concentration averaged over all living levels	From Model #1 (NRRS) and Model #2 (AHS)	Census Tract
L: Average time spent at home	EPA CHAD database	National Average
β: Unit risk, differs by sex and smoking status:	BEIR VI Reference Tables	NA
- Smoking status	From Model#3 (CPS-TUS)	Census Tract
D: Radon daughter decay rate	BEIR/EPA	(constant)
U: Unit conversion factor	BEIR/EPA	(constant)
Y: Expected lifespan, differs by sex and smoking status	Reference Tables (CDC)	National average

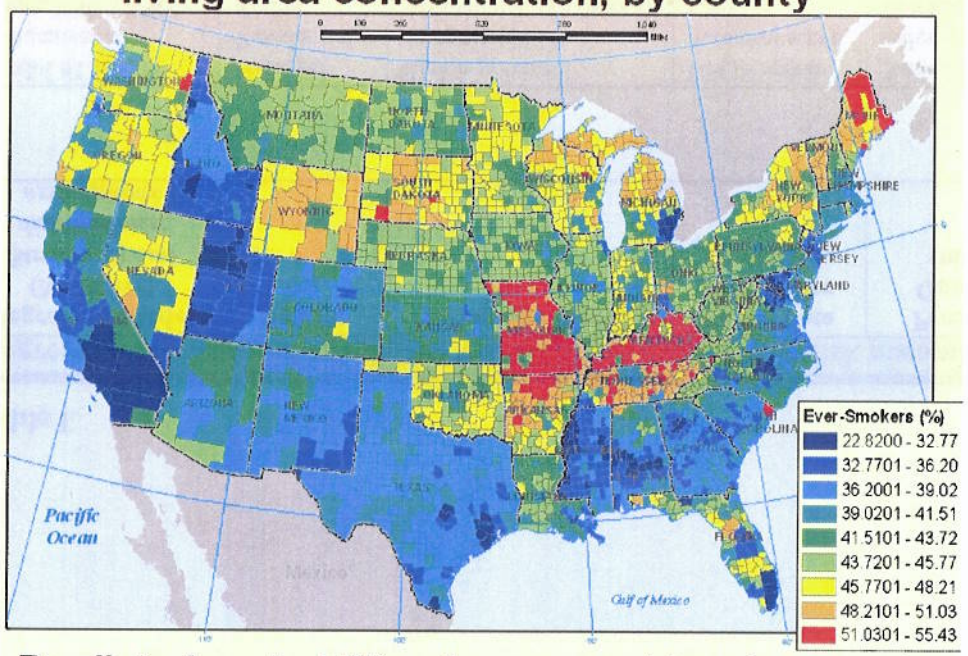
Figure 1. Spatial Patterns of Radon Concentration, Smoking, and Risk It would be inappropriate to use this figure to identify areas as not needing to measure radon concentrations in individual homes nor conclusively identify 'low radon' risks. Even 'lower radon' areas have significant areas of high risk.



Predicted mean annual-average living area concentration, by county



Estimated lifetime risk of fatal lung cancer from residential radon exposure, by county



Predicted probability of ever-smoking, by county

Table II. Log-linear radon concentration model, fixed effects

PARAMETER	ESTIMATE	SE	PVALUE
Intercept	-0.02	0.09	
Northeast	-0.35	0.12	0.003
South	0.026	0.14	0.85
Midwest			
West	-0.18	0.15	0.23
Low Geological Potential			
Medium Geological Potential	0.43	0.074	< 0.001
High Geological Potential	0.74	0.11	< 0.001
Heating Infiltration Degree-Days	0.00006	0.00002	0.008
Average Diurnal Swing	0.041	0.01	< 0.001
Attached Unit	-0.71	0.03	< 0.001
Detached with Basement			
Detached with crawl Space	-0.059	0.04	0.18
Detached with concrete Slab	-0.4283	0.03144	< 0.001
Other Detached	-0.5308	0.1243	< 0.001

Appendix I

Odds Ratios from Multinomial Logistic Regression of House Types, AHS 2007

(Ref grp = single detached houses with basements)

Covariate	Outcome	OR	95%CI	
Poverty vs nopov	attached unit	4.40	4.14	4.66
	crawl space	1.74	1.61	1.88
	slab	1.40	1.30	1.51
	other	2.36	1.91	2.91
Male vs Female	attached unit	0.97	0.94	1.01
	crawl space	1.01	0.96	1.05
	slab	1.00	0.96	1.04
	other	1.04	0.90	1.19
Race: black vs white	attached unit	2.75	2.60	2.92
	crawl space	1.46	1.35	1.57
	slab	1.56	1.46	1.68
	other	1.73	1.40	2.15
Race: asian vs white	attached unit	2.34	2.13	2.56
	crawl space	0.84	0.73	0.97
	slab	2.07	1.86	2.30
	other	2.10	1.48	2.97
Race: other vs white	attached unit	1.76	1.53	2.02
	crawl space	1.22	1.02	1.45
	slab	1.50	1.28	1.76
	other	2.10	1.32	3.33
Race: native vs white	attached unit	1.91	1.59	2.31
	crawl space	1.42	1.14	1.77
	slab	1.74	1.42	2.14
	other	1.83	0.96	3.50
Region: NE vs S	attached unit	0.44	0.42	0.46
	crawl space	0.04	0.03	0.04
	slab	0.03	0.03	0.03
	other	0.09	0.07	0.12
Region: MW vs S	attached unit	0.21	0.20	0.22
	crawl space	0.10	0.09	0.10
	slab	0.04	0.03	0.04
	other	0.12	0.10	0.15
Region: W vs S	attached unit	1.35	1.28	1.43
	crawl space	1.04	0.98	1.10
	slab	1.09	1.03	1.15
	other	0.64	0.53	0.78

Covariate	Outcome	OR	95% CO	
Age: <18 vs 45-54	attached unit	1.11	1.05	1.18
	crawl space	0.87	0.81	0.93
	slab	1.07	1.00	1.14
	other	0.71	0.57	0.89
Age: 18-24 vs 45-54	attached unit	1.99	1.85	2.15
	crawl space	1.01	0.92	1.10
	slab	1.09	1.00	1.19
	other	0.91	0.68	1.24
Age: 25-34 vs 45-54	attached unit	2.68	2.50	2.87
	crawl space	1.17	1.08	1.28
	slab	1.31	1.21	1.42
	other	0.95	0.71	1.26
Age: 35-44 vs 45-54	attached unit	1.25	1.17	1.33
	crawl space	0.94	0.87	1.01
	slab	1.08	1.00	1.16
	other	0.64	0.49	0.84
Age: 55-64 vs 45-54	attached unit	1.08	1.01	1.16
	crawl space	1.08	1.00	1.17
	slab	1.04	0.96	1.13
	other	1.26	0.99	1.61
Age: 65-74 vs 45-54	attached unit	1.16	1.06	1.25
	crawl space	1.19	1.08	1.30
	slab	1.10	1.01	1.21
	other	1.33	1.01	1.76
Age: 75plus vs 45-54	attached unit	1.80	1.66	1.96
	crawl space	1.39	1.26	1.54
	slab	1.10	1.00	1.22
	other	1.71	1.28	2.28

Model Intercepts:	*Estimate*	*(SE)*
Attached unit	1.14	(0.03)
Crawl Space	-0.51	(0.04)
Slab	-0.30	(0.04)
Other	-2.80	(0.10)

Appendix II

Multilevel Logistic Model of ever-smoking in U.S. adults (CPS-TUS 2006-7)

Random Parameters	ESTIMATE	SE			
Variance of state random effect	0.005	0.003			
Variance of random slopes by state for black race	0.035	0.012			
Covariance of random effect for state and random slope for black race	-0.008	0.004			
Variance of CBSA random effect	0.040	0.005			
Variance of household random effect	0.521	0.012			
Fixed Parameters	ESTIMATE	SE	OR	95% CI	
Intercept	*-0.43*	*0.04*			
Male	0.41	0.01	1.51	1.48	1.54
Age 18-24 yrs	-0.86	0.02	0.42	0.41	0.44
Age 25-34 yrs	-0.34	0.02	0.71	0.69	0.73
Age 35-44 yrs	-0.28	0.01	0.76	0.73	0.78
Age 55-64 yrs	0.24	0.02	1.28	1.24	1.32
Age 65-74 yrs	0.33	0.02	1.39	1.34	1.44
Age 75 plus	-0.03	0.02	0.97	0.94	1.01
Poverty	0.37	0.02	1.45	1.41	1.50
Income not reported	-0.24	0.01	0.79	0.77	0.81
Black race	-0.50	0.04	0.61	0.56	0.66
American Indian or Native Alaskan	0.29	0.06	1.33	1.17	1.51
Asian	-1.38	0.05	0.25	0.23	0.28
Native Hawaiian or Pacific Islander	-0.34	0.12	0.71	0.56	0.90
Other / Two or more races	0.20	0.05	1.23	1.11	1.35
Black*Male	0.18	0.03	1.19	1.12	1.28
Native*Male	-0.07	0.09	0.93	0.78	1.11
Asian*Male	0.91	0.06	2.48	2.22	2.78
Islander*Male	0.07	0.17	1.07	0.77	1.48
Other*Male	0.03	0.07	1.03	0.90	1.18
State cigarette excise tax	0.08	0.03	1.08	1.03	1.14
Previous state prevalence (2003)	0.03	0.01	1.03	1.02	1.04
Indoor smoking restrictions in >6 of 7 venue types	0.01	0.03	1.01	0.95	1.08
CBSA % poverty above median	-0.02	0.03	0.98	0.92	1.04
CBSA unidentified/nonmetropolitan	0.04	0.04	1.04	0.97	1.12

REFERENCES

[1] USEPA (2003) EPA assessment of risk from radon in homes. Office of Air and Radiation [EPA 402-R-03-003]

[2] Committee on Improving Risk Analysis Approaches (2009) Science and Decisions: Advancing Risk Assessment. National Research Council of the National Academies, Washington, D.C. Available: http://www.nap.edu/catalog.php?record_id=12209

[3] USEPA (2003) Framework for Cumulative Risk Assessment. Risk Assessment Forum. Washington, DC. EPA-630-P-02-001F.

[4] National Academy of Science (1999) Biological Effects of Ionizing Radiation (BEIR) VI Report: "The Health Effects of Exposure to Indoor Radon." National Academy Press, Washington, D.C.

[5] World Health Organization (2009) WHO Handbook on Indoor Radon: A Public Health Perspective. Available: http://whqlibdoc.who.int/publications/2009/9789241547673_eng.pdf (Accessed 11/22/2010)

[6] Smith, B.J., Field, R.W. (2007) Effect of housing factors and surficial uranium on the spatial prediction of residential radon in Iowa. *Environmetrics* **18**: 481-497

[7] Schumann, R.R., Gundersen, L.C.S., Tanner, A.B. (1994) Geology and Occurance of Radon. *In:* Radon Prevalence, Measurements, Health Risks and Control. Editor: Niren L. Nagda. ASTM International

[8] Greeman, D.J., Rose, A.W. Washington, J.W., Dobos, R.R., Ciolkosz, E.J. (1999) Geochemistry of radium in soils of the Eastern United States. *Applied Geochemistry* **14**: 365-385

[9] Nazaroff W.W. (1988) Predicting the rate of 222Rn entry from soil into the basement of a dwelling due to pressure-driven air flow. *Radiation Protection Dosimetry* **24** 1:199-202

[10] Riley, W.J., Robinson, A.L., Gadgil, A.J., Nazaroff, W.W. (1999) Effects of variable wind speed and direction on radon transport from soil into buildings: model development and exploratory results. *Atmospheric Environment* 33: 2157-2168

[11] Mose, D.G., Mushrush, G.W., Chrosniak, C.E. (1991) A 2-yr study of seasonal indoor radon variations in Northern Virginia. *Journal of Toxicology and Environmental Health* **33**:115-130

[12] United States Geological Survey (1993) National Map of Radon Potential by Geological Province: http://energy.cr.usgs.gov/radon/rnus.html

[13] USEPA (1993) EPA's Map of Radon Zones, National Summary. Office of Air and Radiation, Washington, DC. [EPA 402-R-93-071]

[14] U.S. Geological Survey, Radon Potential State Reports. USGS Earth Science Information Center, Denver, CO. [Open-File Report 93-292]

[15] Phillip N. Price, Anthony Nero, Kenneth Revzan, Michael Apte, Andrew Gelman, W. John Boscardin, et al. High Radon Project Home Page. Lawrence Berkley National Laboratory, Environmental Energy Technologies Division, Indoor Environment Department, High Radon Project. Available: http://eetd.lbl.gov/IEP/high-radon/USgm.htm

[16] Price PN (1997) Predictions and maps of county mean indoor radon concentrations in the mid-Atlantic states. *Health Physics* **72**(6): 893-906

[17] Steck DJ (1992) Spatial and temporal indoor radon variations. *Health Physics* **62**(4): 351-355.

[18] Apte, M.G., Price, P.N., Nero, A.V., Revzan, K.L. (1999) Predicting New Hampshire indoor radon concentrations from geologic information and other covariates. *Environmental Geology* **37**(3): 181

[19] Nero, A.V., Leiden, S.M., Nolan, D.A., Price, P.N., Rein, S., Revzan, K.L., Wollenberg, H.R., Gadgil, A.J. (1994) Statistically based methodologies for mapping of radon actual concentrations – the case of Minnesota. *Radiation Protection Dosimetry* **56**(1-4):215-219

[20] Price, P.N., Nero, A.V., Gelman, A. (1996) Bayesian prediction of mean indoor radon concentrations for Minnesota counties. *Health Physics* **71**(6): 922-936.

[21] Price, P.N., Nero, A.V. (1996) Joint analysis of long- and short-term radon monitoring data from the Northern U.S. *Environment International* **22**:S699-S714.

[22] Datta, G.D., Subramanian, S.V., Colditz, G.A., Kawachi, I., Palmer, J.R., Rosenburg, L. (2006). Individual, neighborhood, and state-level predictors of smoking among U.S. Black women: A multilevel analysis. *Social Science & Medicine* **63**: 2034-1044

[23] Osypuk, T.L., Kawachi, I., Subramanian, S.V., Acevedo-Garcia, D. (2006) Are State Patterns of Smoking Different for Different Racial/Ethnic Groups? An Application of Multilevel Analysis. *Public Health Reports* **121**:563-576.

[24] Shavers, V.L., Faga, P., Jouridine Alexander, L.A., Clayton, R., Doucet, J., Bauzconde-Garbanati, L. (2006) Workplace and home smoking restrictions and racial/ethnic variation in the prevalence and intensity of current cigarette smoking among women by poverty status, TUS-CPS 1998-1999 and 2001-2002. *Journal of Epidemiology in Community Health* **60**(SupplII)ii34-ii43.

[25] Shavers, V.L., Lawrence, D., Fagan, P., Gibson, J.T. (2005) Racial/ethnic variation in cigarette smoking among the civilian U.S. population by occupation and industry, TUS-CPS 1998-1999. *Preventive Medicine* **41**:597-606

[26] National Environmental Justice Advisory Council (2004). Ensuring risk reduction in communities with multiple stressors: environmental justice and cumulative risks/impacts. *Prepared by the National Environmental Justice Advisory Council Cumulative Risks/Impacts Work Group.* Available: http://www.epa.gov/compliance/ej/resources/publications/nejac/nejac-cum-risk-rpt-122104.pdf (Accessed 11/22/2010)

[27] Goldstein, H. (2003) Multilevel Statistical Models. London: Arnold Publishers.

[28] U.S. Census Bureau (2000) Geographic Areas Reference Manual. Available: http://www.census.gov/geo/www/cen_tract.html

[29] Schumann, R., Killgore, M. (1994) The NRRS Soil, Geologic, and Radioactivity Database: Final Report. U.S. Department of the Interior, U.S. Geological Survey, Denver, CO.

[30] Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., et al. (2002). A user's guide to MLwiN. London: Center for Multilevel Modelling, Institute of Education, University of London.

[31] Lucas, R.M., Grillo, R.B., Kemp, S.S. (1992) National Residential Radon Survey Statistical Analysis. Research Triangle Institute, Research Triangle Park, North Carolina [RTI/5158/49-1F]

[32] Apte, M., Nero, A., Revzan, K. (1998) Meteorological Database for the United States. *Indoor Air* **8**:61-67

[33] United States Geological Survey National Map of Radon Potential by Geological Province: http://energy.cr.usgs.gov/radon/rnus.html

[34] U.S. Department of Housing and Urban Development, Office of Policy Development and Research (2009) Codebook for the American Housing Survey, Public Use File: 1997 and later. Available: http://www.huduser.org/Datasets/ahs/AHS_Codebook.pdf

[35] Chahine, T. (2010) Modeling Variability in Environmental Exposures and Health Risks for Community-based Risk Assessment. Doctoral Dissertation, Harvard University, Cambridge, MA

[36] McCurdy T, Glen G, Smith L, Lakkadi Y 2000. The National Exposure Research Laboratory's consolidated human activity database J Expos Anal Environ Epidemiol 10: 566–578. www.epa.gov/chadnet1

[37] Available on: http://apps.nccd.cdc.gov/statesystem/TrendReport/TrendReports.aspx

[38] Lin, C., Gelman, A., Price, P.N., Krantz, D.H. (1999) Analysis of Local Decisions Using Hierarchical Modeling, Applied to Home Radon Measurement and Remediation. *Statistical Science* **14**(3): 305-337. www.stat.columbia.edu/radon