The Links Among Unconventional Oil and Gas Development, Particle Radioactivity and Morality

THE LINKS AMONG UNCONVENTIONAL OIL AND GAS DEVELOPMENT, PARTICLE RADIOACTITIVTY AND MORTALITY

Longxiang Li

A Dissertation Submitted to the Faculty of

The Harvard T.H. Chan School of Public Health

in Partial Fulfillment of the Requirements

for the Degree of Doctor of Science

in the Department of Environmental Health

Harvard University

Boston, Massachusetts.

May 2020

Dissertation Advisor: Dr. Petros Koutrakis Longxiang Li

The Links Among Unconventional Oil and Gas Production, Particle Radioactivity, and Mortality

ABSTRACT

The rapid expansion of unconventional oil and gas development (UOGD) changed the energy market in the past decade, both domestically and globally. Enabled by the technological advancements in hydraulic fracturing and directional drilling, the UOGD could produce crude oil and natural gas from the low-permeability geological formations that were uneconomic to develop in the past. By the end of 2018, carbohydrates extracted from these unconventional formations accounted for 96% and 97% of the domestic crude oil and natural gas production, respectively. Compared with the rapid expansion, our knowledge about its impacts on the environment and residents of neighboring communities is falling far behind. There is an increasing body of literature associating living nearby UOGD sites to adverse health outcomes and harmful environmental exposure. However, current studies are subject to common limitations, including a lack of knowledge regarding exposure pathways, a lack of reliable exposure assessment, and a lack of causal modeling methods applied in investigating the potential associations. This dissertation is designed to address these knowledge gaps partially.

In our first study, we investigated the causal association between exposure to UOGD and all-cause mortality in the Medicare beneficiaries residing in all major production regions of the U.S. We first estimated the association between all-cause mortality and the residential proximity to an active UOGD site; We subsequently divided the participants into two subgroups according to their relative position to the UOGD, a subgroup living upwind to UOGD, and a subgroup living downwind to UOGD, and re-evaluated the subgroup-specific effects. We found that UOGD has significantly greater effects on the residents living downwind, compared to those

living upwind, after adjusting for proximity. Due to the independence of wind direction to residential proximity to UOGD, this association is considered causal.

In our second study, we explored the association of UOGD activity to ambient particle radioactivity, which is a potential exposure pathway of UOGD. Ambient particle radioactivity is the radiometric character of particle matter, which in turn transports the radionuclides from the external to the internal environment, and then expose human tissues to high-energy α particles. We found a significantly positive association between the number of upwind UOGD wells and the gross-β radioactivity measurements monitored by a nationwide operational network.

In our third study, we investigated the association of all-cause mortality with residential exposure to radon, which is the primary source of particle radioactivity. We applied the Difference-in-Difference experimental design, which is a quasi-experimental approach, to investigate the causal linkage in all Medicare beneficiaries residing in New England. We found that a per-unit increase in residential radon exposure is associated with an increment of 4% of all-cause mortality.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This dissertation is a collaborative work, which is impossible without the contribution of co-authors. First, I would like to thank Dr. Petros Koutrakis for his mentorship throughout the five-year doctoral program. I am grateful to each of my committee members for their advice and time. Dr. Brent Coull was an invaluable source of statistical analysis. Dr. Joel Shwartz provided me with great insights into experimental design and answered numerous epidemiological questions. Dr. John Spengler helped me by asking critical questions on behalf of reviewers and readers.

It is my fortune to get into the Environment Exposure & Risk program. Also, it is my honor to get the last Doctor of Science granted by this program. In the past five years, I worked with talented co-authors on each manuscript and sincerely appreciated their collaboration and feedback. Finally, I could not have completed this degree without the love and encouragement of my friends and family.

INTRODUCTION

Unconventional Oil and Gas Development (UOGD) has changed the global energy landscape in the past decade. The large scale adoption of UOGD in the U.S, primarily enabled by the advancements in hydraulic fracturing (commonly referred to as fracking) and directional drilling, converted the U.S from a long-term energy importer to the largest producer of both crude oil and natural gas by the end of 2018. The extensive drilling activities, especially in areas without a significant previous existence of the oil and gas extractive industry, triggered concerns over their potential harmful impacts on the health of residents. Current health effect studies have associated living proximity to UOGD activities with a wide range of harmful outcomes. Simultaneously, there is a growing body of literature associating UOGD with different harmful environmental exposures, shedding light on the potential exposure pathways.

Existing health effects analysis is subject to two significant limitations. First, all existing studies relied on proximity-based exposure assessment proxies, which are vulnerable to be confounded by the social-economic factors. Second, all existing studies associated health outcomes with construction-dependent or operation-dependent indicator, instead of with the chemical or non-chemical agents emitted directly from UOGD. In this thesis, we partially addressed these knowledge gaps by three studies:

AIM I: Exploring the causal association between exposure to UOGD and mortality;

AIM II: Exploring the contribution of UOGD to ambient particle radioactivity;

AIM III: Exploring the association between residential radon exposure and mortality.

AIM II, jointly with AIM III, formed a pathway through which UOGD could influence the health of residents, which is the hypothesis of the first study.

CHAPTER 1: Exposure to Unconventional Oil and Gas Development and All-cause Mortality in Medicare Beneficiaries

Longxiang Li, M.S., Francesca Dominici, Ph.D., Annelise J. Blomberg, Sc.D., Joel D. Schwartz, Ph.D.,

Brent A. Coull, Ph.D., John D. Spengler, Ph.D., Yaguang Wei, M.S., Petros Koutrakis, Ph.D.

Affiliations:

From the Department of Environmental Health (L.L, A.J.B, John D. Spengler, Joel D. Schwartz, B.A.C, W.Y, P.K) and Biostatistics (F.D, B.A.C), Harvard T.H. Chan School of Public Health, Boston.

Abstract

The wide-scale adoption of Unconventional Oil and Gas Development (UOGD) has changed the global energy landscape. However, little is known about whether and how UOGD impact all-cause mortality.

We studied Medicare beneficiaries (N=136,215,059 person-years) in all major UOGD regions in the conterminous U.S. from 2001 to 2015. We obtained records for more than 2.5 million oil and gas wells from Enverus™. For each person-year, we calculated proximity-based exposure (PE) to UOGD and categorized PE into four levels from high to low. To isolate the impact contributed by UOGD-related air pollutants, we calculated the proportion of PE contributed by upwind wells, defined as downwind exposure (DE) to UOGD. Each PE level was dichotomized into downwind and upwind sub-levels (DE^+ and DE^-) accordingly. We used Cox proportional hazard models to estimate the mortality risk associated with each PE level and DE sub-levels. Due to the independence of wind direction on potential confounders, the estimated associations are less vulnerable to unobserved confounding bias.

High PE level was associated with a statistically significant increase in mortality risk compared to the unexposed group (hazard ratio [HR], 1.025; 95% confidence interval [CI], 1.021 to 1.029). Within high PE, the HR associated with DE^+ is 1.031 (95% CI 1.025 to 1.037), significantly higher than that associated with DE^- (HR 1.022, 95% CI, 1.016 to 1.027), when both are compared to the same unexposed group.

Introduction

Oil and natural gas development from the low-permeability unconventional geological formation (known as unconventional oil and gas development [UOGD]) has rapidly expanded over the past decade. As of 2015, over 100,000 onshore UOGD wells had been drilled via a practice involving directional drilling combined with multistage high-volume hydraulic fracturing (fracking).(U.S.

Energy Information Administration (EIA) 2019c) As of 2015, 17.6 million U.S. residents live within one kilometer of at least one active well.(Czolowski et al. 2017) Although, UOGD is typically more productive than conventional oil and gas development (COGD) in extracting hydrocarbons, it generally has more extended construction periods and requires larger water volumes, proppants and chemicals during the multi-stage fracking process resulting in larger waste volumes.(U.S. Enverionmental Preotection Agency (EPA) 2016)

Proximity to UOGD has been associated with increased human exposure to both chemical and non-chemical agents. (Landrigan, Frumkin, and Lundberg 2020) UOGD-related air contaminants include volatile organic carbons (VOCs) (Allen 2014) nitrogen oxides (Cheadle et al. 2017) and naturally occurring radioactive materials. (Casey et al. 2015) UOGD operations were also associated with elevated concentrations of organic compounds,(Hill and Ma, n.d.) chloride, and total suspending solids in drinking water. (Olmstead et al. 2013) Elevated UOGD-related agents (noise(Blair et al. 2018) and night light(Franklin et al. 2019)) have been reported in nearby neighborhoods. Based on these findings, proximity-based exposure metrics (PEs) were designed as surrogates of UOGD exposure and applied in retrospective epidemiologic studies finding a significant association between UOGD exposure with adverse prenatal,(Casey et al. 2016) respiratory,(Rasmussen et al. 2016; Koehler et al. 2018) cardiovascular, and carcinogenic outcomes. (McKenzie et al. 2019)

The link between UOGD exposure and all-cause mortality has not been investigated on a national scale. Existing studies regarding UOGD health effects are based on PEs which assume a uniform and isotropic (irrespective of spatial directionality) declining impact of UOGD as the distance between UOGD and subjects increases.(Health Effects Institute-Energy (HEI-Energy) Research Committee 2019b) These PEs fail to account for different transport pathways of

UOGD-related agents in environmental media. Existing studies based solely on PE cannot distinguish contributions from different exposure pathways. UOGD also has significant socio-cultural impacts on host communities. (Kelsey, Partridge, and White 2016; Perry 2013) Associations found by previous observational studies based on PEs were vulnerable to be biased by unobserved confounding health determinants that are potentially influenced by UOGD.

We conducted the most extensive study to date on the health effects of UOGD exposure to address these knowledge gaps. We hypothesized that mortality risks are higher for communities in proximity to, and downwind of, UOGD operations due to the role of air transport. We first calculated the PE for each person-year at risk and then calculated the proportion of PE contributed by wells located upwind of a given location (downwind exposure [DE]). Two survival models were fitted to associate UOGD exposure with all-cause mortality via PE only or via a PE and DE combination. Atmospheric movement is theoretically independent of potential confounders such as age, gender, and socioeconomic status (SES). The DE-based survival analysis created a quasi-experiment design to better control both measured and unmeasured confounding bias. (Dominici, Greenstone, and Sunstein 2014)

Methods

Mortality Data

Our study area (Figure 1-1) includes all ZIP Codes within or around seven major shales defined by the U.S. Energy Information Administration (Figure 1-S1).(U.S. Energy Information Administration (EIA) 2019a) We grouped these ZIP Codes into three non-adjacent areas: northern, eastern, and southern subregions. The Medicare beneficiaries denominator file was obtained from the Center for Medicare and Medicaid Services.(ResDac 2018) We built an open

cohort with person-years of follow-up for Medicare beneficiaries 65 or older at enrollment and residing in the study area for at least one year during 2001-2015. For each person-year of follow-up at risk, we extracted details including age, race, sex, Medicaid eligibility, residential ZIP Code, and date of death.

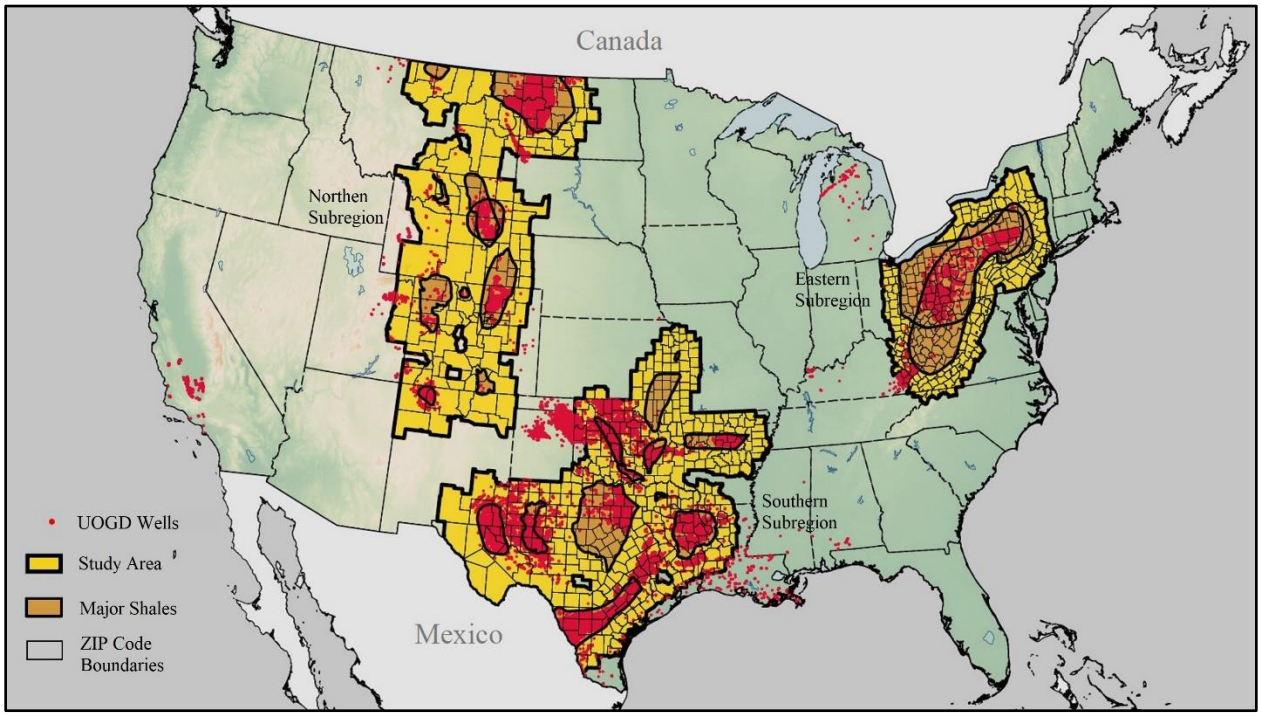


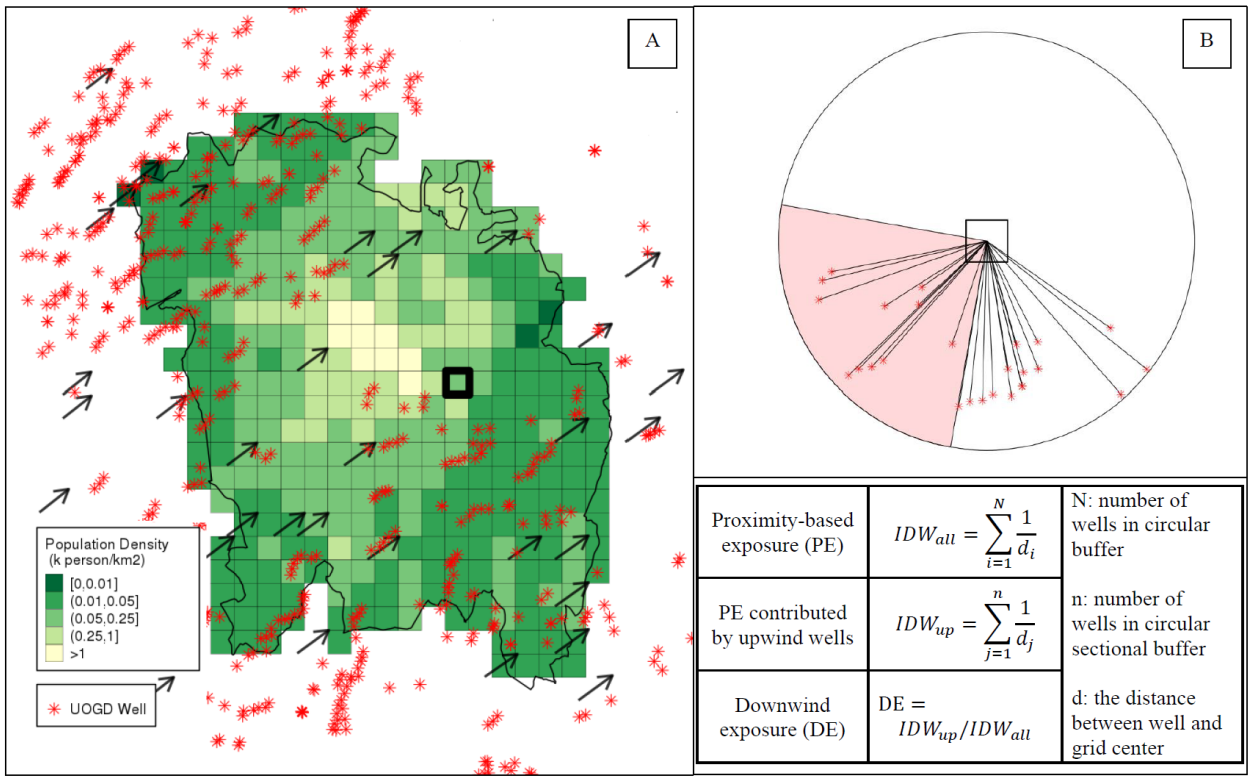
Figure 1-1. Map of the study area, which contains over 120,000 active UOGD wells located in 9244 ZIP Codes in December 2015.

UOGD Data

We obtained information on domestic well location, construction, and production from Enverus™ (formerly Drillinginfo.com) through its Direct Access Service.(Enverus 2019) Additional information on Enverus™ is presented in Supplementary Appendix (SA)(Page. 18). We categorized horizontally-drilled wells as UOGD, and vertically-drilled wells as COGD.(U.S. Enverionmental Preotection Agency (EPA) 2016) Wells under construction or in production are considered active.

Exposure Assessment

We used an inverse-distance-weighting (IDW) method to calculate monthly PE to UOGD. This metric incorporates the distances from UOGD wells and the number of UOGD wells nearby. We calculated distances between the grid center and each active well within a 15-km circular buffer in each month for each 1-km grid cell (Figure 1-2A). We then calculated grid-specific PE by summing the inverse of these distances (Figure 1-2B). This method gave higher weight to UOGD closer to the grid center. The spatial resolution of residential location information is ZIP Code-level in the Medicare cohort. As a result, we calculated ZIP Code-level PE by taking a weighted average of grid-level PE according to grid-level population density.(Doxsey-Whitfield et al. 2015)



Proximity-based exposure (PE)	$IDW_{all} = \sum_{i=1}^{N} \frac{1}{d_i}$	N: number of wells in circular buffer
PE contributed by upwind wells	$IDW_{up} = \sum_{j=1}^{n} \frac{1}{d_j}$	n: number of wells in circular sectional buffer
Downwind exposure (DE)	$DE = IDW_{up}/IDW_{all}$	d: the distance between well and grid center

Figure 1-2. UOGD exposure assessment in an example ZIP Code and month (Washington, PA, 15301, August 2015).

Communities downwind of UOGD wells are more likely to be affected by the air pollutants emitted on-site and transported by air. To characterize this phenomenon, we also calculated a

ZIP Code-level monthly DE ranging from 0 to 100%. First, a grid-specific monthly prevailing wind was estimated by downscaling the monthly wind field provided in North America Regional Reanalysis (NARR) by bilinear spline interpolation.(Mesinger et al. 2006) We calculated the monthly proportion of PE contributed by upwind wells, defined as wells within the windward circular sectional quadrant whose central angle is 90 degrees (Figure 2B). For example, if wells were evenly distributed within the circular buffer, then one-quarter of wells would fall within the windward quadrant, and the upwind contribution would be 25%. These grid-specific DE were also aggregated to ZIP Code-level weighted by population and averaged by year.

Potential confounding variables

To control for potential confounding bias, we considered the following individual-level covariates available in the Medicare denominator files: gender (male or female), race (White or Black), age, calendar year, and Medicaid eligibility (yes or no) as a proxy for low socioeconomic status. We considered ZIP Code Tabulation Area (ZCTA)-level indicators of socioeconomic status including annual median household income, owner-occupied housing units median value, population percent below the poverty line, population percent without high school diplomas, population density, and homeownership rate. These were obtained from the 2000 and 2010 US Census and the American Community Survey (ACS) and linearly extrapolated to account for the covariates' time-varying nature. County-level covariates, including annual percent of non-smokers and obese people, were obtained from the Behavioral Risk Factor Surveillance System (BRFSS).(CDC (Center for Disease Control and Prevention) 2013)

Some air pollutants emitted by UOGD-related operations can also originate from other sources, such as traffic and COGD activities. To isolate specific UOGD effects, we included multiple

environmental covariates to control for non-UOGD sources. COGD exposure was calculated using the same approach used to generate UOGD exposure metrics. We also controlled for $PM_{2.5}$ exposures from other anthropogenic activities using gridded annual ambient levels of $PM_{2.5}$ predicted by a previously published national spatiotemporal model.(Di et al. 2016) We obtained annual land cover data from the U.S. Geological Survey(Earth Resources Observation and Science (EROS) Center 2012) and calculated the ZIP Code-specific percent of land surface covered by vegetation and developed area to represent the residential emission of pollutants other than $PM_{2.5}$.

Statistical Analysis

We fitted two extended Cox proportional models with time-dependent covariates (Anderson-Gill model) to investigate the health effects of UOGD exposure.(Andersen and Gill 1982) Robust sandwich variance estimators were used to account for the nesting of observations within ZIP Code in both models.(Lee et al. 1992) We allowed baseline mortality rate to vary by gender, race, eligibility for Medicaid, age categories and calendar year categories in both models (SA Section 3). Model I relied on PE, estimating the health effects of living proximity to UGOD. Model II was jointly based on PE and DE, and able to estimate the influence of UOGD on downwind communities specifically.

In Model I, person-years with zero PE exposure were used as the unexposed group (reference level). Most of UOGD activities are within the shales included in this study. As a result, person-years outside of our study area due to beneficiaries' mobility were merged into the unexposed group. All person-years with non-zero PE were categorized by quartiles into four exposure groups: low (0, 25^{th} percentile], medium-low (25^{th}, 50^{th} percentile], medium-high (50^{th}, 75^{th}

percentile] and high PE group (75^{th}, 100^{th} percentile]. We used subgroup analysis to evaluate specific populations effects and subregional analysis to evaluate consistency across three subregions. We also tested the sensitivity of these associations to breaking point selection by trying other categorization methods. Finally, we assessed the sensitivity of Model I to the inclusion and exclusion of covariates by refitting modified Model I with a subset of covariates.

Model II was fitted to investigate the difference in mortality effect estimates between downwind and upwind exposure to UOGD, holding the proximity constant. Toward this end, each PE level was divided into DE sub-level (upwind contribution of PE $\geq$ 25%, indicating the population in a ZIP Code is predominately downwind of wells, DE^{+}) and upwind sub-level (upwind contribution < 25%, indicating the population in a ZIP Code is not predominately downwind of wells, DE^{-}). Four PE groups in Model I were divided into eight exposure groups of Model II, four of which are for DE^{+}, the other four are for DE^{-}. The unexposed group in Model I was not subdivided. This formulation was equivalent to adding an interaction term between PE and DE to the formula in Model I (SA Section 3). We also performed subgroup, subregional and sensitivity analysis of Model II using similar approaches to those used for Model I.

The analysis was conducted on the Cannon cluster, supported by the Research Computing Group, and on the Research Computing Environment, supported by the Institute for Quantitative Social Science, both at Harvard University, Faculty of Arts and Sciences. We used R software (version 3.4.2)(R Core Team 2017) and survival package (version 3.1.8)(Therneau 2019) to perform survival analysis.

Results

Table 1-1. Characteristics of population grouped by PE and DE to UOGD.

Covariates	No-PE*	Low-PE		High-PE		ASD†
	--	DE$^-$	DE$^+$	DE$^-$	DE$^+$	DE$^+$ vs DE$^-$
No. of Beneficiaries	13,176,937	1,221,558	790,820	753,219	717,874	--
No. of Person-years	110,093,570	4,321,587	2,505,366	3,236,837	3,025,157	--
No. of Deaths	5,434,451	214,430	127,428	154,964	143,965	--
Mortality (%) ‡	4.9	5.0	5.1	4.8	4.8	--
Individual level						
Female (%) §	57.6	56.4	56.5	56.1	56.3	0.004
White (%)	90.7	91.7	92.6	89.6	88.4	0.007
Medicaid Eligibility (%)	10	12.2	12.9	11	10.8	0.009
Age (year)	75.1±7.6	74.8±7.5	74.9±7.5	74.6±7.5	74.6±7.5	0.005
ZIP Code-level						
PE of COGD	5.0±14.0	12.0±19.9	12.9±19.4	15.4±25.7	13.7±22.9	0.007
$PM_{2.5}$ ($\mu g/m^3$)	10.3±2.7	8.9±2.6	9.1±2.4	9.3±1.6	9.1±1.4	0.028
Below Poverty (%)	18.1±4.1	20.8±4.9	20.8±4.9	23.0±4.0	23.3±4.0	0.018
No High School (%)	8.2±5.8	9.7±5.5	10.1±5.6	8.8±5.8	8.6±6.2	0.041
Pop Density (100 person/km^2)	25.0±13.8	25.7±13.2	27.7±14.2	22.3±12.9	21.6±12.4	0.007
Mean Household Income ($\times 10^3$ \$)	17.7±25.9	4.6± 8.8	3.5± 7.0	10.8±13.9	13.9±16.6	0.011
Median House Value ($\times 10^5$ \$)	51.8±20.9	46.8±16.8	44.1±12.3	52.9±21.1	53.7±20.4	0.033
County-level						
BMI ¶	27.0±1.0	27.4±1.2	27.5±1.3	27.4±1.2	27.4±1.0	0.043
Non-Smoker (%)	46.9±6.5	46.7±8.0	47.6±7.8	43.9±7.4	45.4±6.6	0.136

This cohort included 136,215,059 person-years from 15,198,496 subjects living in 9244 ZIP Code areas from 2001 to 2015. A total of 174,624 UOGD wells were completed during our study period (2001-2015) which increased PE in all three subregions (Table 1-S1and Figure 1-S3). The high PE group had a lower percentage of Medicaid eligibility and poverty rate; and higher average median income and median house value than the other three PE groups. This aligns with a previous study that found a higher SES in communities near UOGD (Table 1-1).(Boudet et al. 2018) We found the absolute standardized difference (ASD) of each potential confounder between DE^+ and DE^- subgroups within each PE group is always lower than 10% (Table 1-1 and Figure 1-S2). This suggests that confounding bias is negligible.

According to Model I all PE levels were associated with statistically significant mortality risk increases, compared to the unexposed level (Table 1-S3). The associated risks of mortality increase monotonically when the PE level increases from low to high (Figure 1-3A). High PE level was associated with a statistically significantly elevated risk of all-cause mortality (HR 1.025; 95% CI, 1.021 to 1.029). According to the subgroup analysis, the estimated mortality risk associated with the female subgroup is greater than the male subgroup within each PE level (Figure 1-S5A). We did not find evidence of racial or age-dependent difference (Page 29).

According to Model II, living downwind of UOGD wells was associated with a higher risk of death compared to living upwind of UOGD (Figure 3B, Table S3). More specifically, within the high PE level, we found evidence of a statistically significant increase in the risk of death when comparing the DE^+ subgroup with the unexposed group (HR 1.031; 95% CI, 1.025 to 1.037). When comparing the DE^- subgroup to the unexposed group, the associated HR is 1.022 (95% CI, 1.017 to 1.028), significantly lower than the HR associated with the downwind subgroup (95% CI, 0.003 to 0.014, p-value<0.001). The significant downwind-upwind difference in the

increased risk of mortality persisted in the medium-high and medium-low PE levels (Table 1-S3). According to the subgroup analysis, the downwind-upwind difference is more remarkable in male and younger subgroups compared to female and elder subgroups (Figure 1-S5).

When we stratified the analysis by subregions, we found similar associations between UOGD exposure, both PE and DE, and all-cause mortality (Figure 4). The effects of UOGD exposure on mortality in the northern and eastern subregions were larger than those in the southern subregion. The associations found in Model I and Model II were robust to the breaking point selection in categorization (Page 29).

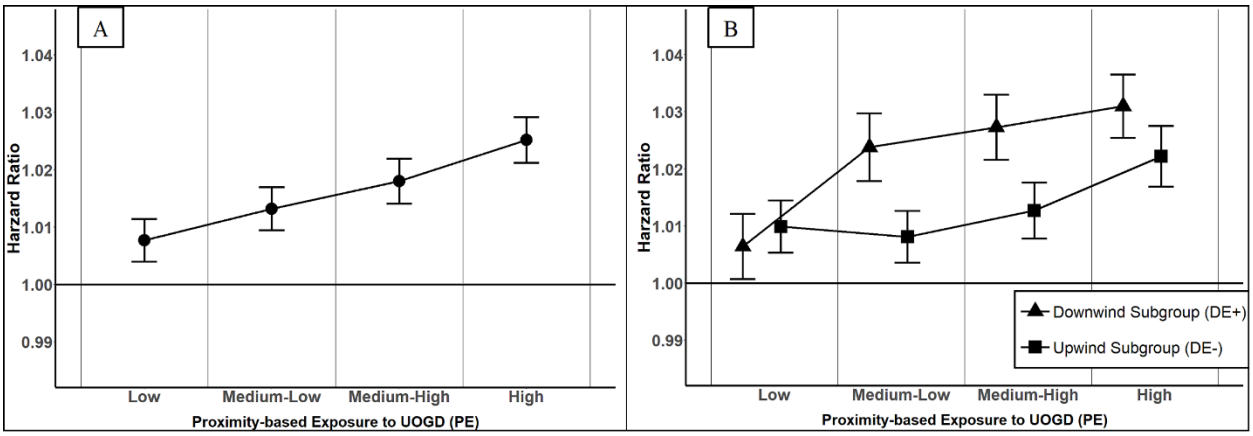


Figure 1-3. The estimated relative risk of mortality associated with each level of proximity-based exposure to UOGD (PE) and both sub-levels of downwind exposure to UOGD (DE) within each PE level.

Panel A shows the result of Model I, which investigates whether the relative risk of mortality associated with each PE level when compared to the no-exposure group is significantly higher than 1. Panel B shows the result of Model II. Model II first investigated the association between PE and all-cause mortality in DE^+ groups and DE^- groups, respectively. We can compare the relative risks associated with the DE^+ subgroup and DE^- subgroup within each PE levels.

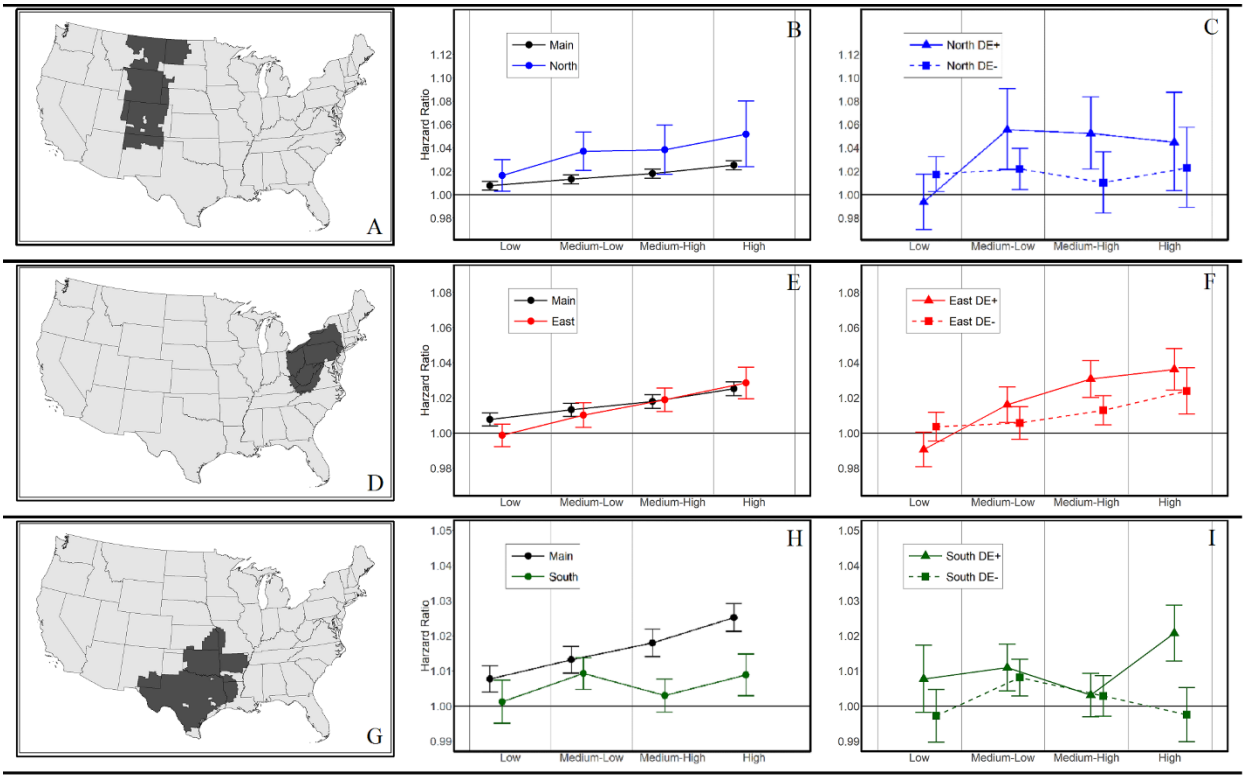


Figure 1-4. The relative risk associated with exposure to UOGD in each subregion.

Panel A shows the extent of northern subregion; Panels B and C show the result of two corresponding models respectively; Panel D shows the extent of eastern subregion; Panels E and F show the result of two corresponding models; Panel G shows the extent of southern; Panels H and I show the result of two corresponding models.

Discussion

We found that residential exposure to UOGD, characterized by PE and DE exposure metrics, was associated with a significantly elevated risk of mortality in Medicare beneficiaries. This was observed in all three subregions (Figure 1-3), both genders, all age groups, and major races (Page 29). These findings indicate that the past decade of extensive expansion of onshore UOGD has impacted the health of nearby communities regardless of geological, environmental, or demographic factors.

Communities surrounding UOGD are exposed to diverse chemical and physical pollutants. We recognized that proximity-based exposure, such as the PE we constructed, assumed a uniform

distance-decay gradient of UOGD exposure and does not account for transport mechanisms. To address this, we leveraged the quasi-random variation of wind direction to isolate the impact contributed by agents transported by air. We found pronounced difference in the mortality risks associated with DE^{+} and DE^{-} sublevels within three PE levels (Table 1-S3). These results indicate that airborne contaminants emitted at UOGD and transported downwind contribute to the increased mortality. We also found statistically significant, but lower relative risk, for populations residing upwind of UOGD wells. These associations could be due to other agents whose transport are independent of atmospheric movement, such as surface and groundwater contaminants, traffic-dependent impacts, noise, light pollution, and lifestyle disruption. They could also be explained by UOGD-related airborne pollutants transported to upwind communities but at a lower frequency than downwind communities.

Previous studies on adverse health outcomes of UOGD exposure were challenged by the Health Effects Institute in its recently-released literature review for relying on exposure surrogates instead of utilizing specific measurements of UOGD-related contaminants.(Health Effects Institute-Energy (HEI-Energy) Research Committee 2019b) However, airborne pollutants emitted by UOGD can originate from other sources, making it difficult to disentangle the specific contribution of UOGD to human health.(Health Effects Institute-Energy (HEI-Energy) Research Committee 2019a) Currently operating population-oriented monitor networks are not designed to characterize neighborhood-scale gradients of UOGD emitted airborne pollutants. As a result, exposure surrogates are essential for large-scale population-based studies even though direct measurements of exposures may be possible in smaller cohort studies. Our novel methodology specifically assesses the contribution of UOGD-related airborne pollution and may be used to investigate the health effects of other industrial activities with complex, difficult-to-

measure pollution profiles.

A key strength of our study is the application of wind direction to create a quasi-experiment (detailed in Model II). For previous observational UOGD health effects studies, a primary challenge was the study population could not be randomly assigned to exposed and nonexposed groups, suggesting that people who live closer to UOGD have different initial health conditions (likely due to different SES) from those further distances from UOGD operations. Observational studies based solely on proximity-based exposure metrics are vulnerable to unobserved confounding factors bias – a key limitation in existing studies.(Health Effects Institute-Energy (HEI-Energy) Research Committee 2019b) In this study, the assignment of DE sub-levels was determined by wind direction whose variation is theoretically independent of the potential confounding factors such as SES (SA Section 4). This quasi-experiment design could mitigate confounding and omitted covariates bias in the estimated association.(Dominici, Greenstone, and Sunstein 2014)

Our study also used a nationwide cohort of over 15 million Medicare beneficiaries, starting prior to UOGD expansion, and a comprehensive database covering over 2.5 million oil and gas wells. Medicare beneficiaries include over 95% of U.S. citizens 65 or older.(Schwartz et al. 2018) Our study population is nationally-representative and has a low influence of occupational UOGD exposure. The comprehensive geographic coverage, encompassing all shales, allowed us to analyze nationally and regionally associations. Previous regional studies have not examined the regional heterogeneity of observed associations. The subregional analysis could be regarded as three independently conducted epidemiology studies involving the association based on an identical exposure-outcome pair.

A limitation of our study is that we were unable to estimate associations between mortality and specific UOGD-related airborne agent(s), due to unavailability of high-resolution exposure data of air pollutants other than $PM_{2.5}$. Instead, our study estimates the effects of an air pollutants mix originating from UOGD wells. Further observational studies near UOGD, especially in both wind directions, are necessary to identify air pollutant(s) responsible for the health effects observed in this study. We were not able account for well characteristics including drilling depth, product type, well age, productivity, operator, and wastewater management method in our exposure assessment due to a lack of information for all domestic major shales. These factors could contribute to between-well variations in emission intensity and thus could help further improve the exposure metrics.

Considering the increased rate and scale of the expansion of UOGD, it is critical to understand the potential health risks associated with this industry. In this study, we designed two metrics and employed them to estimate the health effects of living close to, and downwind of, UOGD respectively. According to our models, we conclude that residential exposure to UOGD is positively associated with an elevated risk of all-cause mortality in Medicare population.

Appendix: Supplementary Material

Appendix 1.1: Details of study extent

Our study extent (shown in Figure 1-1) covers all counties around or within major UOGD production regions. A total of 9244 ZIP Codes within these counties were included. The target formations (shales) underground these regions include Bakken formation primarily underneath western North Dakota; Niobrara formation primarily underneath eastern Colorado and northern Utah; Marcellus and Utica formation underneath Pennsylvania, West Virginia, eastern Ohio and southern New York; Woodford formation underneath Oklahoma; Fayetteville formation underneath central Arkansas; Permian formation underneath southeastern New Mexico and western Texas; Barnett formation underneath central and northern Texas; Eagle Ford formation underneath southern Texas and Haynesville formation underneath eastern Texas and western Louisiana.

The boundaries of this formation are determined by EIA and shown as dark yellow polygons. Most UOGD wells completed by the end of 2015 (shown as red dots in Figure 1-1) are positioned to this extent. Counties around Bakken shale and Niobrara shale are clustered as a northern subregion. Counties around Marcellus shale and Utica shale are clustered as an eastern subregion. All other counties were grouped as a southern subregion.

Appendix 1.2: Data source of exposure assessment

State energy agencies are the primary data source of Enverus™. Wells in Enverus™ database are classified into horizontal wells, vertical wells, and directional wells. However, the definition of directional wells varies among states. For example, directionally drilled wells and horizontally

drilled wells are both grouped in a single class in Colorado as directional well, while are kept separated in New Mexico. As a result, there is a visually remarkable difference in the percentage of horizontal wells, which is considered unconventional wells in our study, across the state line even though they share the target geological formation and apply similar drilling methods.

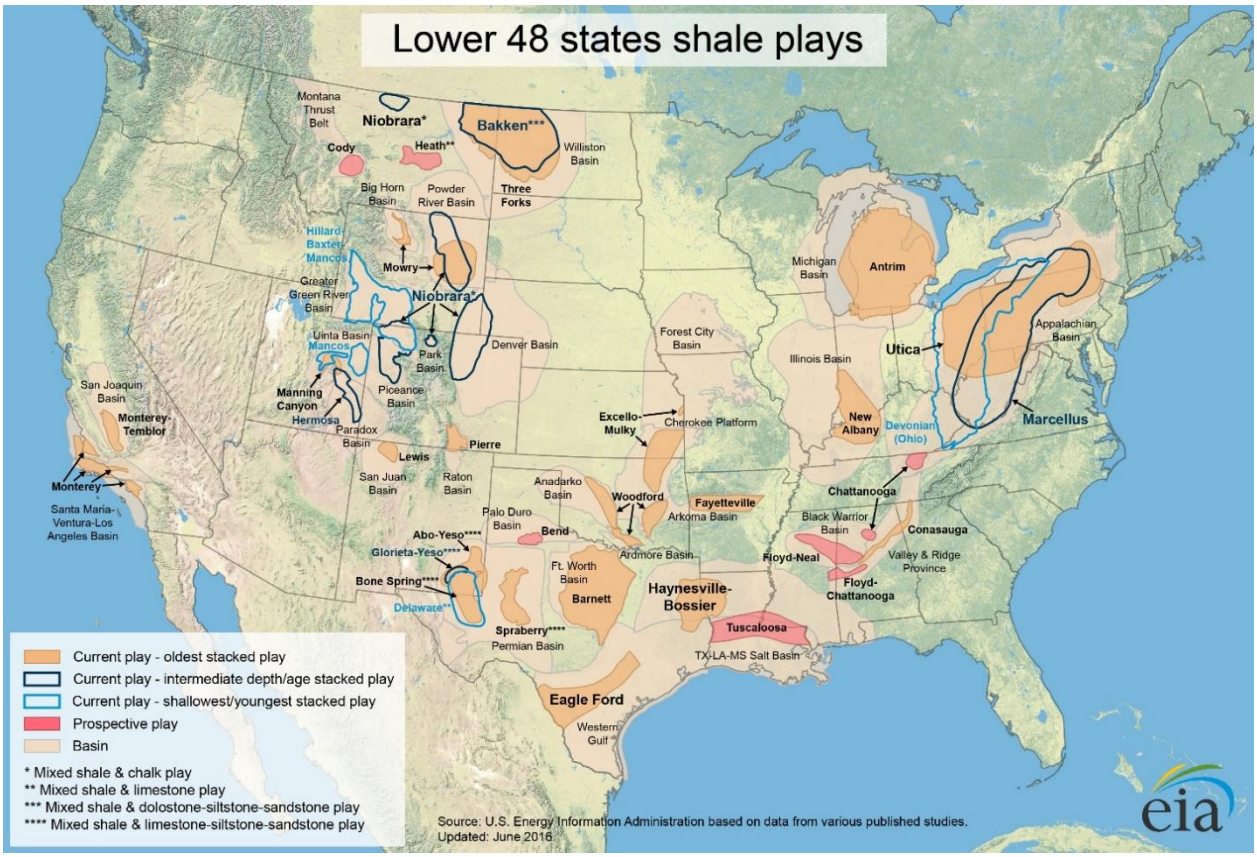


Figure 1-S1.The extent of major UOGD production regions defined by the US. Energy Information Administration, a subsidiary of the U.S Department of Energy.

Consequently, it is not reliable to assume that all directionally drilled wells are UOGD. Also, the raw dataset from Enverus™ does not provide drilling type information for more than 75% of the wells, mostly drilled before 2000. It is also inaccurate to assume that all wells without drilling

type information as conventional because some state agencies do not require drilling type. To solve these problems, we need to predict the binary drilling type based on the drilling types of its neighbor wells and other secondary information.

We trained a random forest model to perform this binary prediction. Random forest is a regression tree-based algorithm good at capturing the non-linear relation between the primary variable and secondary variables, thus suitable for solving this binary classification problem. Secondary variables incorporated in the model include: 1) the drilling type of the nearest well with known drilling type information; 2) the distance to the nearest conventional/unconventional well; 3) the percentage of conventional and unconventional wells of the nearest 10 wells with known drilling type; 4) the O&G reservoir where the well is positioned; 5) the spudding/completion time; 6) the drilling depth; 7) the natural gas /liquid production in the first 6 months; 8) the production declining rate of gas/liquid. After running a grid search for optimal performance, the parameters of this model were set as the following: the number of trees is 100, maximum depth is 15, the minimum size of a node is 5. The accuracy of this model is 99.83% for conventional O&G wells and 93.1% for unconventional O&G wells. The performance difference is potentially caused by the re-fracturing process of some conventional O&G wells. As shown in Figure S2, the covariates with the three highest importance are the number of conventional wells within 10 km, the number of unconventional wells within 10 km, the drilling type of the closest well with known drilling type. We used the random forest algorithm implemented in h2o to fit and evaluate our model.

We excluded wells without any temporal information about spudding, completion, production, and abandonment. We then calculated the annual region-specific average length of the construction period and estimated the construction-related dates for wells with only production

records.

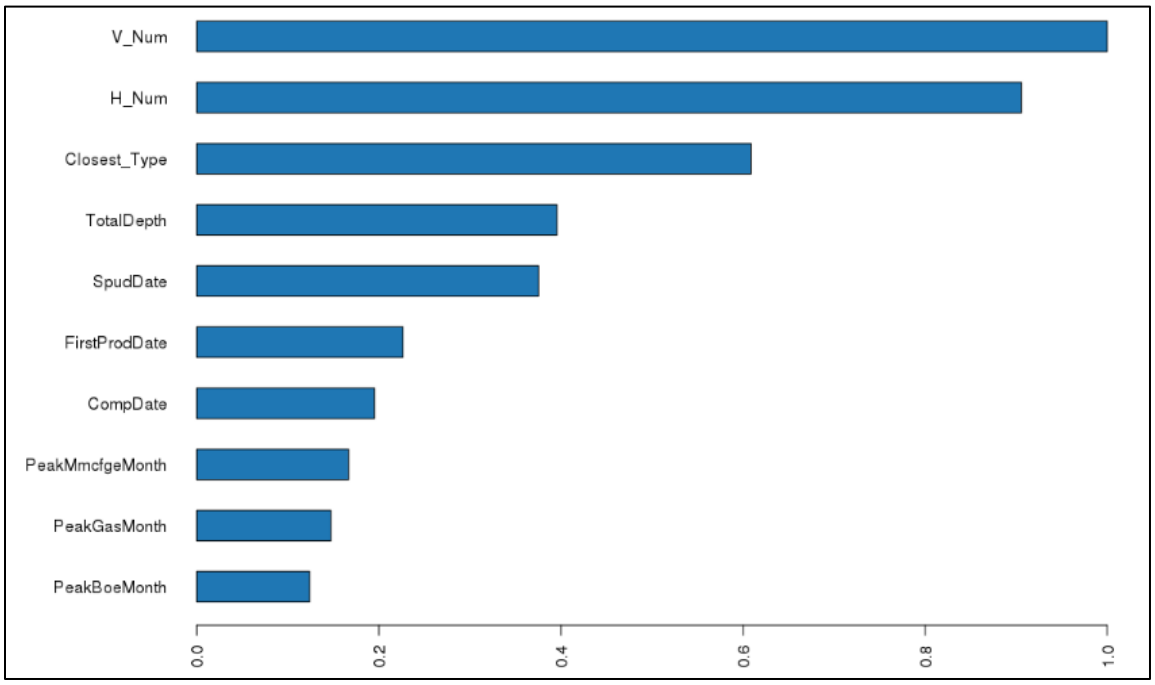


Figure 1-S2.The relative importance of covariates in the random forest model.

Full names of the variables from top to bottom: Number of conventional wells within 10km, number of unconventional wells within 10 km; the drilling type of the closest well with known drilling type. total drilling depth, spudding date, the date of the first production record, completion date, the date of peak million cubic feet of gas equivalent production, the date of peak natural gas production, the date of peak barrel of oil equivalent production.

Appendix 1.3: Details about statistical analysis

In Model I, we used the following formula:

$$\lambda\left(t\middle|\boldsymbol{Z}(t)\right) = \lambda_0 exp\left(\boldsymbol{\beta} \times \boldsymbol{Z}(t)\right)$$

$\boldsymbol{Z}(t) =$	$\boldsymbol{PE}(t) +$	Exposure term of Model I
	$\text{strata}\big(Gender, Race, Medicaid, Year, Age(t)\big) +$	Individual-level factors
	$PM_{2.5}(t) + \%Develop(t) + \%Veg(t) + COGD(t) +$	ZIP Code-level environmental factors
	$\%Highschool(t) + \%Poverty(t) + Density(t) +$ $Housevalue(t) + Income(t) + \%Ownership(t) +$	ZIP Code-level socioeconomic factors
	$\%Smoking(t) + BMI(t)$	County-level behavior risk factors

where $\boldsymbol{Z}(t)$ is a mix of time-varying covariates, including the exposure term, individual-level demographic factors, ZIP Code-level environmental factors, ZIP Code-level socioeconomic factors, and county-level behavior risk factors. The hazard at time t (denoted as $\lambda(t)$) depends on the value of covariates at that time (denoted as $\boldsymbol{Z}(t)$). The regression coefficient of $\boldsymbol{Z}(\cdot)$, denoted as $\boldsymbol{\beta}$, is constant over time. The stratification term $strata\big(Gender, Race, Medicaid, Year, Age(t)\big)$ allows the baseline function, denoted as λ_0To vary across subgroups. The exposure term of Model I is only PE (denoted as $\boldsymbol{PE}(t)$).

We added an interaction term between downwind-based exposure to UOGD (DE) and proximity-based exposure to UOGD (PE) to the formula of Model I. In this way, we had a modified $\boldsymbol{Z}^*(t)$ as

$\boldsymbol{Z}^*(t) =$	$\boldsymbol{DE}(t) \times \boldsymbol{PE}(t) + \boldsymbol{PE}(t)$	Exposure term of Model II
	$\text{strata}\big(Gender, Race, Medicaid, Year, Age(t)\big) +$	Individual-level factors
	$PM_{2.5}(t) + \%Develop(t) + \%Veg(t) + COGD(t) +$	ZIP Code-level environmental factors
	$\%Highschool(t) + \%Poverty(t) + Density(t) +$ $Housevalue(t) + Income(t) + \%Ownership(t) +$	ZIP Code-level socioeconomic factors
	$\%Smoking(t) + BMI(t)$	County-level behavior risk factors

Appendix 1.4: Balance of covariates between exposure groups

As shown in Table 1-1, the absolute standardized difference (ASD) of the individual- and ZIP Code-level covariates are all below 10% except for the county-level smoking rate, indicating an overall balanced assignment of DE^+ and DE^- sublevels. We also calculated ASD between DE+ and DE- sublevels within each PE levels (Figure 1-S3.The absolute standardized difference of all covariates between DE+ and DE- sub-levels within each PE level.). Individual-level covariates, including race, age, gender, eligibility for Medicaid, were well balanced in each PE levels because all ASD were smaller than 10%. ZIP Code- and county-level covariates were not as well balanced. However, most of the ASDs were below 10%, with a few exceptions over 20%, indicating a general balanced assignment.

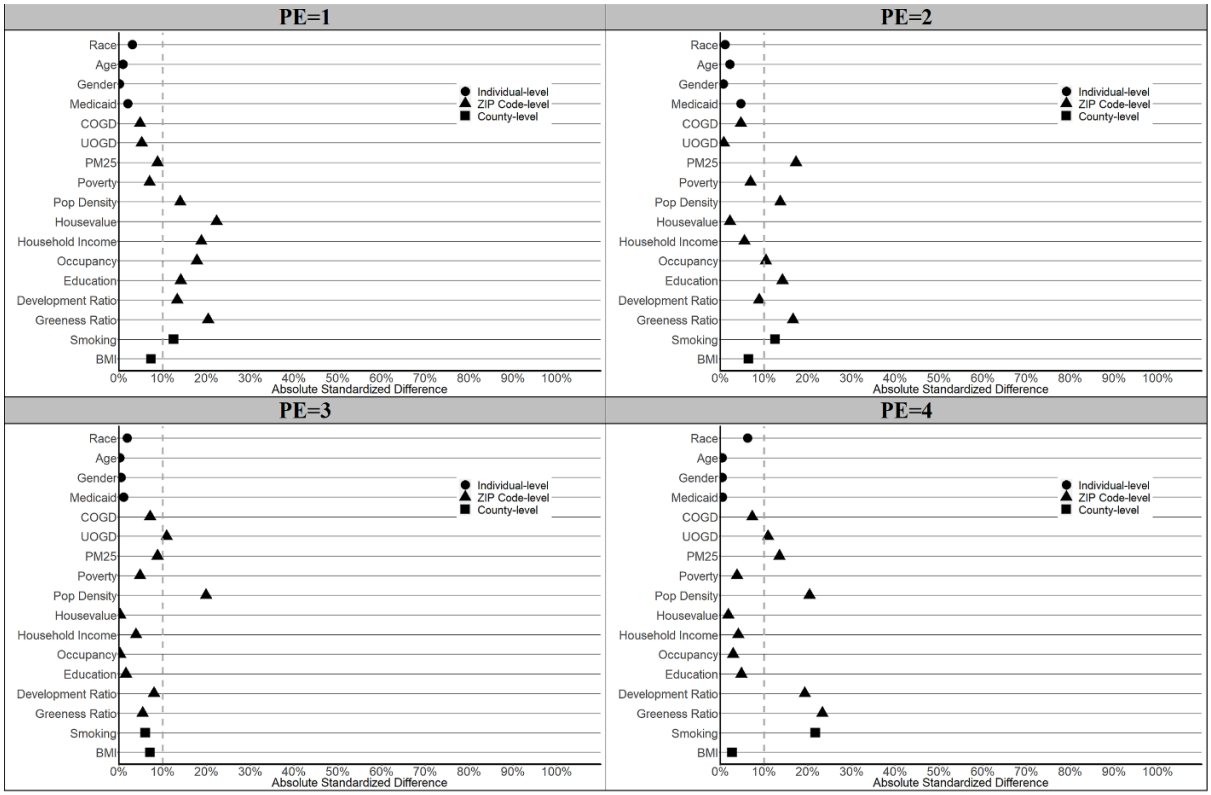


Figure 1-S3.The absolute standardized difference of all covariates between DE+ and DE- sub-levels within each PE level.

Appendix 1.5: Spatio-temporal patterns of UOGD and COGD exposure

During our study period, UOGD expanded rapidly across major shale regions (Figure 1-S4). We calculated the average PE levels for both UOGD and COGD in each subregion during our study period (Figure 1-S4). Due to the rapid expansion of UOGD, the mean of PE to UOGD increased, especially after 2007 in southern subregion and after 2011 in northern and eastern subregions. Meanwhile, the expansion of COGD was not as rapid. The COGD exposure even declined slightly in southern subregion. At the beginning of our study period (Figure 1-S4 (A)), UOGD was not widely adopted for economic reasons. Only two small areas, one above Eagle Ford shale and one above Bakken shale, were in high PE level. Approximately 40,000 residents lived in these regions. At then end of our study period (Figure 1-S4 (B)), over 18 million U.S residents were living within the high PE to UOGD level regions across conterminous U.S.

Table 1-S1. The spatio-temporal variability of PE to UOGD and PE to COGD during 2001-2015

Years	PE to UOGD	PE to COGD	Person-Years	Deaths	Mortality
Northern Region					
[2001,2003]	0.021	2.920	1,761,323	87,398	4.96%
[2004,2007]	0.032	3.370	2,524,946	115,638	4.58%
[2008,2011]	0.071	3.400	2,849,780	122,281	4.29%
[2012,2015]	0.403	3.700	3,256,531	132,032	4.05%
Eastern Region					
[2001,2003]	0.001	7.340	14,662,750	791,712	5.40%
[2004,2007]	0.002	8.100	19,642,761	1,015,742	5.17%
[2008,2011]	0.065	8.920	20,397,279	1,004,922	4.93%
[2012,2015]	0.417	9.080	21,635,110	1,032,606	4.77%
Southern Region					
[2001,2003]	0.235	4.620	8,739,250	474,872	5.43%
[2004,2007]	0.656	5.040	12,341,594	627,732	5.09%
[2008,2011]	3.100	5.450	13,540,645	644,641	4.76%
[2012,2015]	4.890	5.350	14,863,090	682,075	4.59%

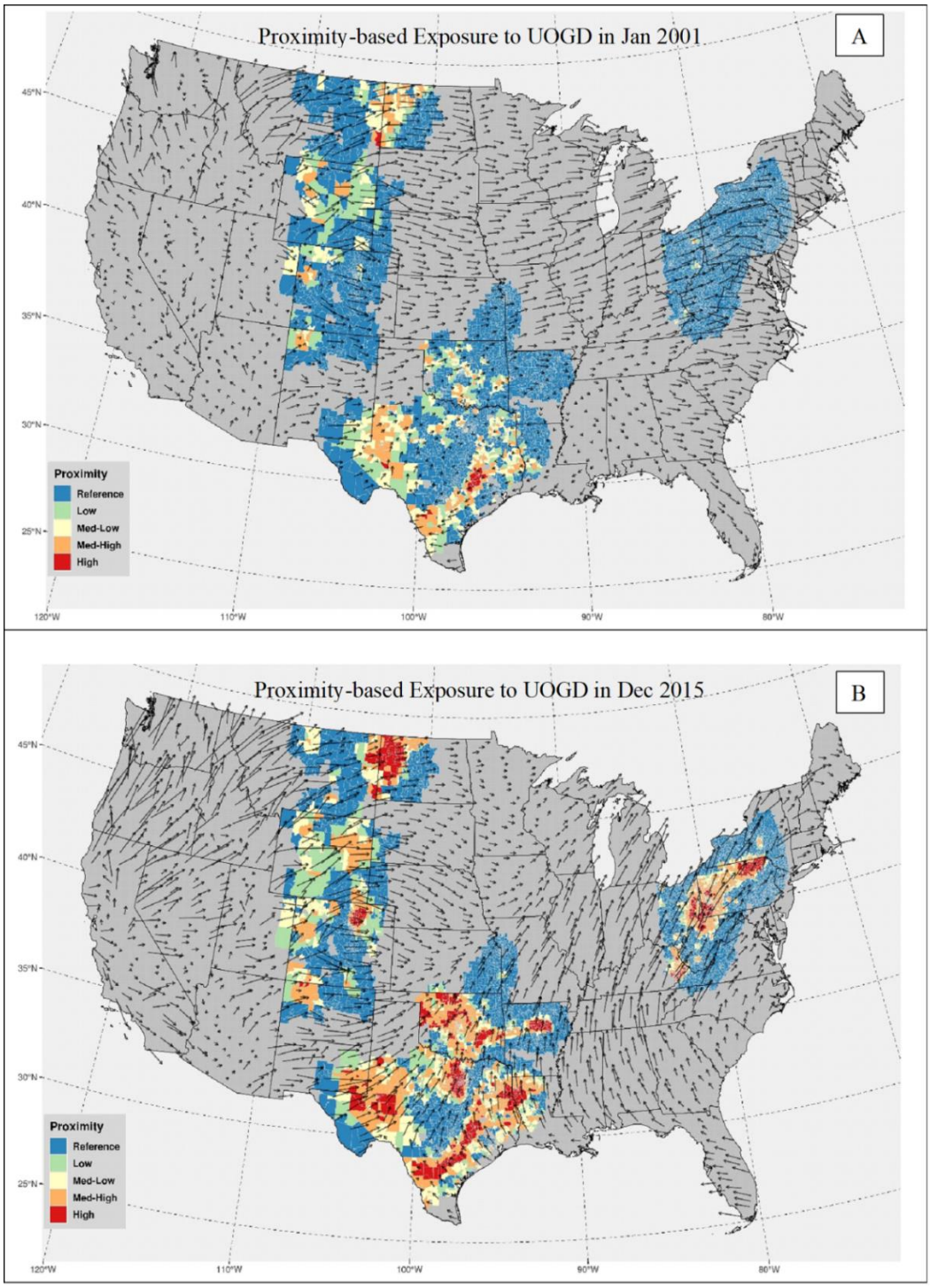


Figure 1-S4. The comparison between PE to UOGD in the beginning (Panel A) and at the end (Panel B) of our study period.

Appendix 1.6: Spatio-temporal patterns of mortality

We calculated the raw mortality (ratio between the number of events and the number of person-years at risk of event) for each subregion and year range pair (Table 1-S2). The mortality declined gradually in each PE levels due to advances in health care during our study period. Higher PE to UOGD groups has higher mortality than lower PE to UOGD groups besides few exceptions.

Table 1-S2. The spatio-temporal variation of all-cause mortality in study population. Mortality is reported as the ratio between number of event and number of person-year at risk

	Reference	Low	Medium Low	Medium High	High
Northern Subregion					
2001-2003	77,983/1,575,652 4.95%	3,591/69,160 5.19%	3,888/79,830 4.87%	1,936/36,681 5.28%	0/0 NA
2004-2007	102,672/2,253,425 4.56%	4,358/93,009 4.69%	5,840/122,545 4.77%	2,701/54,730 4.94%	67/1,237 5.42%
2008-2011	104,579/2,449,975 4.27%	6,188/139,326 4.44%	8,030/185,782 4.32%	3,088/67,022 4.61%	396/7,675 5.16%
2012-2015	98,960/2,449,937 4.04%	10,344/240,617 4.30%	12,335/320,012 3.85%	8,247/200,242 4.12%	2,146/45,723 4.69%
Eastern Subregion					
2001-2003	782,079/14,486,115 5.40%	3,466/66,636 5.20%	3,285/59,185 5.55%	2,882/50,814 5.67%	0/0 NA
2004-2007	990,334/19,175,492 5.16%	7,671/143,847 5.33%	12,016/219,839 5.47%	5,721/103,583 5.52%	0/0 NA
2008-2011	895,354/18,280,148 4.90%	23,285/451,796 5.15%	51,096/999,132 5.11%	32,904/623,530 5.28%	2,283/42,673 5.35%
2012-2015	825,715/17,527,948 4.71%	20,883/418,662 4.99%	71,607/1,425,377 5.02%	98,211/1,939,811 5.06%	16,190/323,312 5.01%
Southern Subregion					
2001-2003	359,408/6,643,702 5.41%	24,759/446,503 5.55%	57,969/1,048,056 5.53%	28,613/523,220 5.47%	4,123/77,769 5.30%
2004-2007	416,667/8,262,887 5.04%	33,677/654,522 5.15%	82,203/1,553,188 5.29%	77,392/1,504,525 5.14%	17,793/366,472 4.86%
2008-2011	385,067/8,204,443 4.69%	37,033/772,909 4.79%	79,931/1,602,337 4.99%	86,138/1,730,883 4.98%	56,472/1,230,073 4.59%
2012-2015	395,633/8,783,846 4.50%	30,365/643,499 4.72%	82,546/1,739,660 4.74%	100,088/2,056,407 4.87%	73,443/1,639,678 4.48%

Appendix 1.7: Subgroup analysis

We performed subgroup analysis by restricting person-years to a specific subgroup of population and refitting both Model I and Model II. As shown in Figure 1-S5(A), relative risks associated with PE levels in female enrollee are uniformly higher than those of male enrollee. As shown in Figure 1-S5(B), the the relative risks associated with DE+ subgroups are higher than those associated with the corresponding DE- subgroups, except for in the low PE level, regardless of the gender of the beneficiaries. However, the downwind-upwind difference is more pronounced in male beneficiareis compared to female beneficiaries. This difference could be explained by the gender-based behavior distinction. As shown in Figure 1-S5(C) and Figure 1-S5 (D), the estimated risks associated with each PE level in African American subgroups have wider CI than those for white beneficiaries, due to a smaller sample size. For this reason, most of the exposure levels are not significantly associated with elevated risk, thus making it difficult to investigate the ethnic-based difference in the association. As shown in Figure 1-S5 (E) and Figure 1-S5G), there is no remarkable difference in the PE-associated risks of mortality among four age groups. However, the downwind vs upwind difference in the associated risk is more pronounced in the younger two age subgroups (a subgroup from 65 to 75 and a subgroup from 76 to 85), compared to the elder two age subgroups (a subgroup from 86 to 95 and a subgroup from 96 to 105). These age-dependent difference in the associated risk could also be explained by the behaviro distinction. For example, younger medicare beneficiaries are more active and thus spend more time outdoors compared to their senior counterparts.

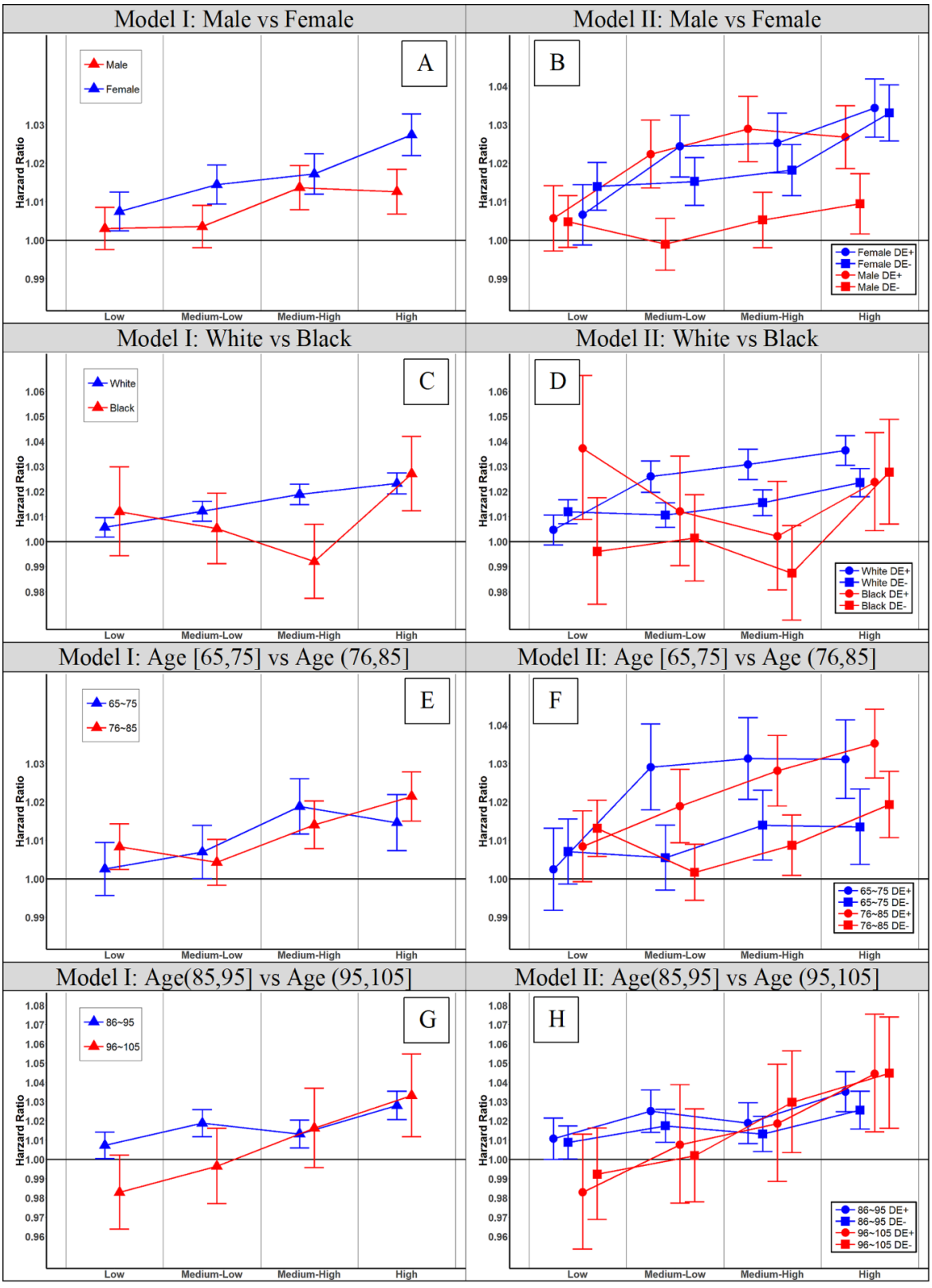


Figure 1-S5.Results of the subgroup analysis of both models.

Appendix 1.8: Results of Model I and Model II

Table 1-S3 The results of full Model I and Model II (the data source of Figure 1-3in the main text)

PE level	DE level	Hazard Ratio	95% CI	P-value	DE+ vs DE- Diff	95%CI	P-value
Low	--	1.008	(1.004,1.011)	<0.001	--	--	--
Med-Low	--	1.013	(1.009,1.017)	<0.001	--	--	--
Med-High	--	1.018	(1.014,1.022)	<0.001	--	--	--
High	--	1.025	(1.021,1.029)	<0.001	--	--	--
Low	DE-	1.010	(1.005,1.014)	<0.001	--	--	--
Low	DE+	1.006	(1.001,1.012)	0.0142	-0.004	(-0.009, 0.002)	0.884
Med-Low	DE-	1.008	(1.004,1.013)	<0.001	--	--	--
Med-Low	DE+	1.024	(1.018,1.030)	<0.001	0.016	(0.010,0.022)	<0.001
Med-High	DE-	1.013	(1.008,1.018)	<0.001	--	--	--
Med-High	DE+	1.027	(1.022,1.033)	<0.001	0.015	(0.009,0.020)	<0.001
High	DE-	1.022	(1.017,1.028)	<0.001	--	--	--
High	DE+	1.031	(1.025,1.037)	<0.001	0.009	(0.003,0.014)	<0.001

Appendix 1.9: Robustness to the exclusion/inclusion of covariates

Basic Model I and Model II do not account for any ZIP Code- or county-level covariates. According to the results of basic Model I (Figure 1-S6A), every PE level is significantly associated with an elevated risk of mortality. But the estimated risk is not monotonically increasing along with a higher PE level. In addition, the estimated risks are remarkably higher

than those predicted by full Model I. However, the downwind-upwind difference in the relative risk estimated by basic Model II is pronounced for three PE levels (Figure 1-S6B).

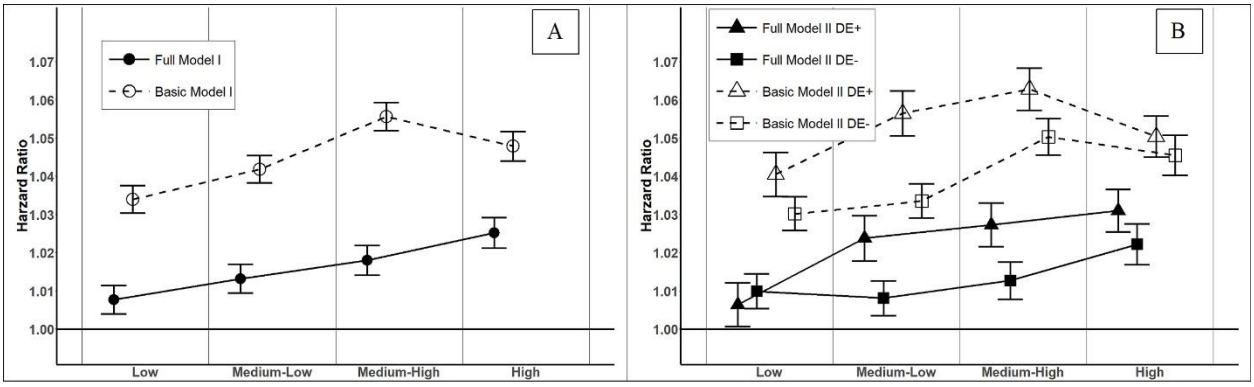


Figure 1-S6.A comparison between the full models and the basic models in which only individual-level factors are adjusted.

Panel A shows the comparsion of a full Model I and a correpsonding basic Model I without adjusting for ZIP Code-level environmental factors, ZIP Code-level SES factor or county-level behavior risk; Panel B shows the comparison of a full Model II and a basic Model II without adjusting for ZIP Code-level environmental factors, ZIP Code-level SES factor or county-level behavior risk.

Moderately simplified Model I and Model II only adjust for individual-level covariates and ZIP Code-level environmental factors. According to the moderately simplified Model I (Figure 1-S8A), the association between PE to UOGD and all-cause mortality is still not monotonic. But the estimated relative risks are in the same magnitude with those from full Model I. According to the moderately simplified Model II (Figure 1-S8B), the downwind-upwind difference in the estimated risks is remarkable in all PE levels, not influenced by omitting ZIP Code-level covariates and county-level behavior risk factors.

Lightly simplified Model I and Model II adjust for individual factors, ZIP Code-level environmental factors and ZIP Code-level SES covariates while omitting county-level behavior risk factors. As shown in Figure 1-S8(A), the lightly simplified Model I estimated lower relative

risk for each level of PE. However, omitting county-level behavior risk factor does not change the trend of the association. As shown in Figure 1-S8(B), the downwind-upwind difference in the estimated risk is robust to omitting county-level behavior risk factor.

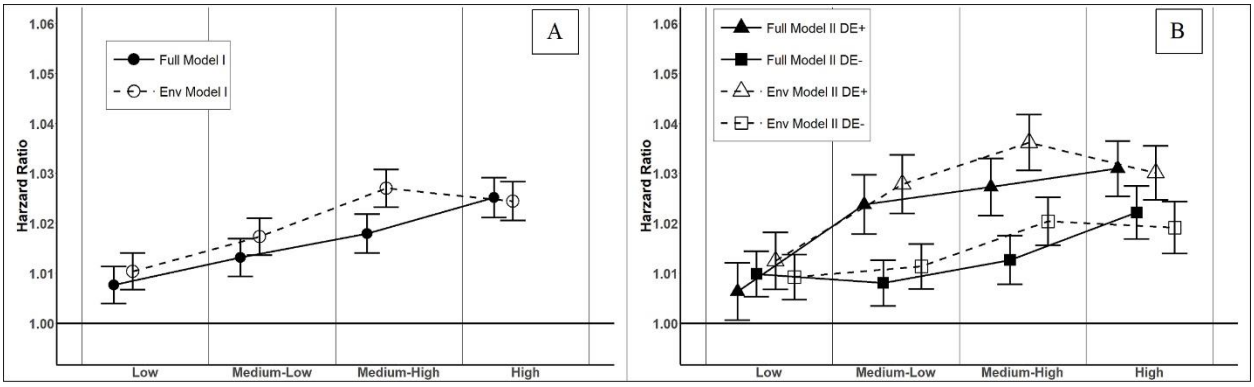


Figure 1-S7 A comparison between full models and the moderately simplified models in which only individual-level factor, ZIP Code-level environmental factors are adjusted.

Panel A shows the comparsion of a full Model I and a correpsonding moderately simplified Model I without adjusting for ZIP Code-level SES factor or county-level behavior risk; Panel B shows the comparison of a full Model II and a correspongind lightly modified Model II without adjusting for ZIP Code-level SES factor or county-level behavior risk.

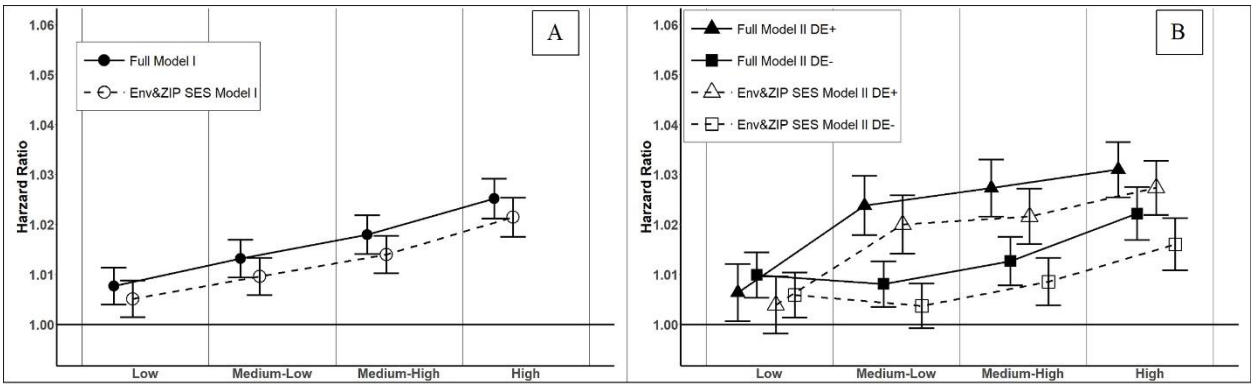


Figure 1-S8.A comparison between full models and the lightly simplied models in which only individual-level factor, ZIP Code-level environmental factor and social econocmical facotor are adjusted.

Panel A shows the comparsion of a full Model I and a correpsonding lightly modified Model I without adjusting for county-level behavior risk; Panel B shows the comparison of a full Model II and a correspongind lightly modified Model II without adjusting for county-level behavior risk.

Then we focus on the influence of omitting COGD exposure terms which is highlighted in HEI's review as a key limitation is some of the previously published studies. We fitted a single-pollutant Model I which only contains PE to UOGD without adjusting for PE for COGD. As shown in Figure 1-S9(A), excluding COGD exposure term from the formula of Model I leads to a higher predicted risk associated with each UOGD PE level. This indicates the necessity to include COGD exposure in health effect studies concerning UOGD. Models unadjusted for COGD may overestimate the health effects of UOGD. Despite predicting greater risk, the results of the single-pollutant Model I still show a clear trend that greater PE to UOGD is associated with a greater risk of all-cause mortality (Figure 1-S9(A)). According to the results of single-pollutant Model II (Figure 1-S9(B)), the downwind-upwind difference is robust to the omitting of COGD.

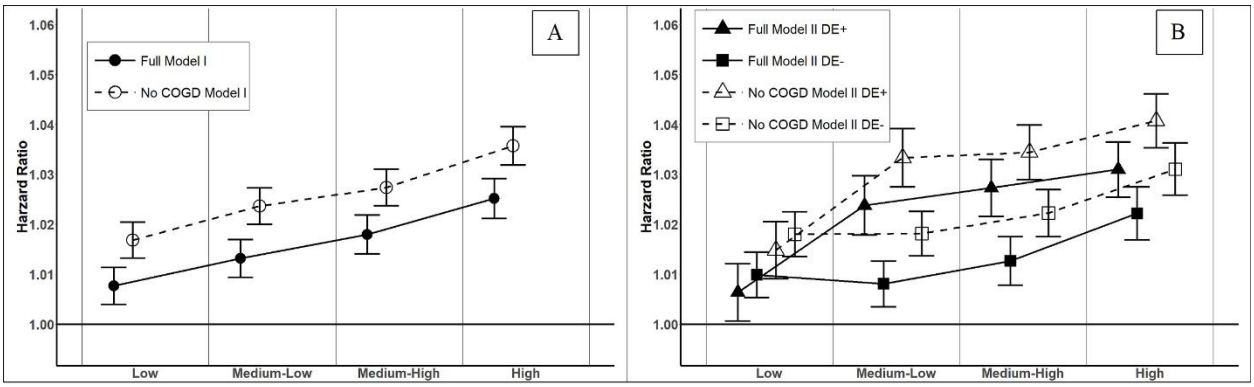


Figure 1-S9.A comparison between full models and the modified models in which PE to COGD is omitted.

Panel A shows the comparsion of a full Model I and a correpsonding modified Model I without adjusting for COGD; Panel B shows the comparison of a full Model II and a correspongind modified Model II without adjusting for COGD.

Appendix 1.10: Robustness to breaking points selection

In the original analysis, we categorized above-zero continuous PE to UOGD by its quartiles into four levels, low level (0, 25th percentile], medium-low level (25th ,50th percentile], medium-high level (50th, 75th percentile] and high level (75th ,100th percentile]. The estimated relative risk associated with each PE level are shown in Figure S10 as bold black cross. We designed another four sets of percentile-based cutting points: Set 1 (15th, 50th and 85th percentiles), Set 2 (20th, 50th and 80th percentiles), Set 3 (30th, 50th, and 70th percentiles) and Set 4 (35th, 50th and 65th percentiles). We re-categorized the PE based on these four sets of cutting points and re-fitted both models respectively to test the robustness of the associations to the cutting points selection.

The relative risks of mortality estimated by these four models are comparable because they are compared to the same reference group of no exposure. We can plot them together in Figure S10. Each estimated risk is visualized by a black cross whose horizontal line represents the range determined by the cutting points, whose vertical line represents the 95% CI of the estimated risk, whose center point is located in the mid-point of the percentile range. As shown in Figure 1-S10, the estimated mortality risk is in general linearly associated with the mid-point of PE range, indicating that the results of Model I is robust to cutting point selection.

After re-categorizing populations based on four sets of cutting points, we further dichonomized the new PE-based subgroups into downwind and upwind subgroups in the same method with the original analysis. According to the results of the re-fitted Model IIs (Figure 1-S11), the downwind-upwind difference in the estimated mortality risk is pronounced in all medium-low and medium-high PE levels. When the high PE level starts higher than the 75th percentile (> 80th percentile in Set 2 and >85th percentile in Set 1), the downwind-upwind difference is

pronounced. When the high PE level starts lower than 75th percentile (>65th percentile in Set 3 and >60th percentile in Set 2), this difference is not as remarkable as the first two sets.

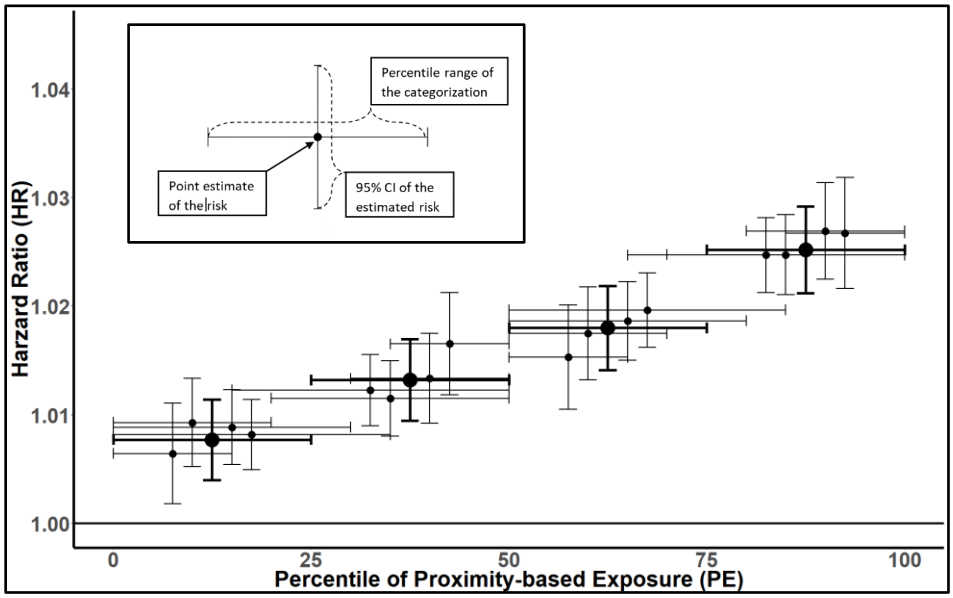


Figure 1-S10. The estimated risk of mortality associated with each category of PE to UOGD according to the sensitivity anslysis.

Bold black cross indicate the results of original Model I.

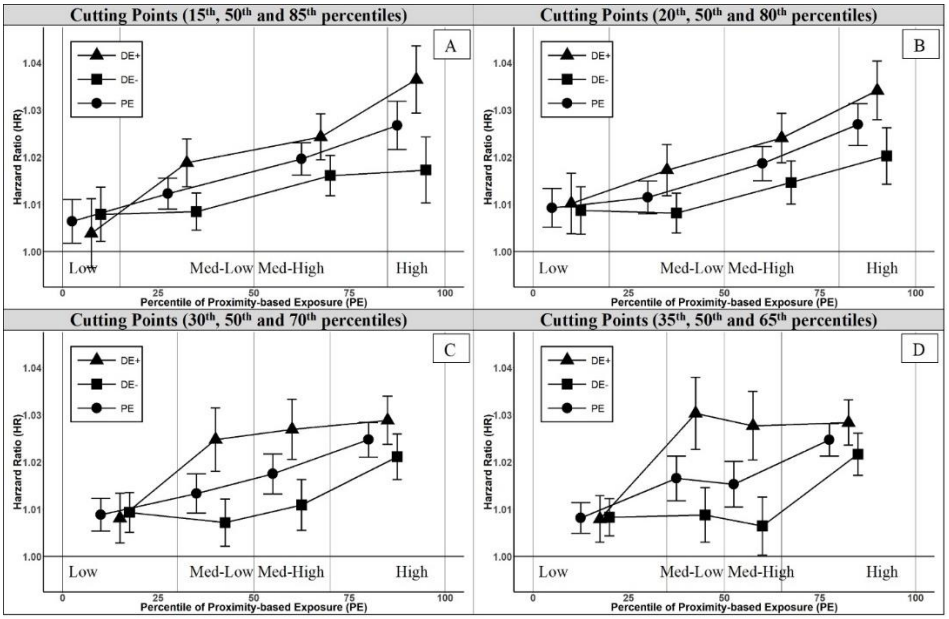


Figure 1-S11. The estimated risk of mortality associated with each category of PE to UOGD

CHAPTER 2: Unconventional Oil and Gas Development Raises Airborne Particle Radioactivity

Longxiang Li, M.S, Annelise J. Blomberg, Sc.D., Joel D. Schwartz, Ph.D., Brent A. Coull, Ph.D., John D. Spengler, Ph.D., Petros Koutrakis, Ph.D.

Affiliations:

From the Department of Environmental Health (L.L, A.J.B, John D. Spengler, Joel D. Schwartz, B.A.C, P.K) and Biostatistics (B.A.C), Harvard T.H. Chan School of Public Health, Boston.

Abstract

Unconventional Oil and Gas Development (UOGD) expanded extensively in the United States from the early 2000s. However, the influence of UOGD on the radioactivity of ambient particulate is not well understood. We used the ambient particle radioactivity (PR) measurements carried out by RadNet, an operational nationwide environmental radiation monitoring network. We then obtained the location and production information of over 1.5 million wells from the Enverus database. We investigated the association between the upwind UOGD well count and the downwind gross-beta radiation with the adjustment for environmental factors governing the natural emission and transport of radioactivity. Our statistical analysis found that an additional 100 upwind UOGD wells within 20 km is associated with an increase of 0.024 mBq/m^3 in the gross-beta particle radiation downwind from the wells. Based on the published health analysis of PR, the widespread UOGD could induce adverse health effects to residents living close to UOGD by elevating PR.

Introduction

The extraction of crude oil and natural gas from the low-permeability unconventional geological accumulating formation (known as unconventional oil and gas development [UOGD]) expanded extensively over the past decade. As of 2017, over 120,000 onshore UOGD wells had been drilled via a practice involving directional drilling combined with multistage high-volume hydraulic fracturing (fracking).(U.S. Energy Information Administration (EIA) 2019c) Meanwhile, numerous controversies have arisen, partially due to the potential harmful impacts on the local environment,(Allen 2014; Cheadle et al. 2017; Hill and Ma, n.d.; Olmstead et al. 2013; Blair et al. 2018; Franklin et al. 2019) and on the health of nearby residents.(Casey et al.

2016; Rasmussen et al. 2016; Koehler et al. 2018; McKenzie et al. 2019)

Naturally occurring radioactive material (NORM) is a common by-product in Oil and Gas (O&G) production industry. The concentration of Uranium-238 in sedimentary formation rich in organic matter, such as black shale, is significantly higher than the background level of in earth's crust due to the natural attenuation process.(Commission 1961; Cordeiro et al. 2016) Before widespread UOGD, studies had detected above-background levels of Radium-226, a decay product of U-238, in the wastes of conventional oil and gas development (COGD).(Kolb and Wojcik 1985; Fisher 1998) Regarding UOGD, enhanced levels of U-238 and Ra-226 have recently been detected in the produced water from unconventional hydrocarbon reservoirs,(Torres, Yadav, and Khan 2018; Brown 2014) in the drill cuttings from the lateral drilling within the unconventional formation,(Eitrheim et al. 2016; Pennsylvania Department of Environmental Protection 2016; Zhang, Hammack, and Vidic 2015) in the impoundment sediments,(Rich and Crosby 2013) in the soil of brine spill accident scene,(Lauer, Harkness, and Vengosh 2016) and in the stream sediments near discharging sites.(Lauer, Warner, and Vengosh 2018) Two studies in the Marcellus shale region found a positive association between UOGD activities and indoor levels of Radon-222, a gaseous decay product of Ra-226.(Casey et al. 2015; Xu, Sajja, and Kumar 2019)

However, the influence of UOGD on the radioactivity of ambient particles (referred to as particle radioactivity [PR]) is not well understood. The particle-bound progeny of Radon-222 (referred to as radon in this study) contribute to the majority of PR.(Hernández et al. 2005; Baskaran 2011) Radon firstly decays into a chain of short-lived particle-reactive progeny. These short-lived radionuclides quickly react with the water molecules and atmospheric gases passing by, form ultrafine clusters and finally attach to airborne particles.(Porstendörfer 1994; Gründel and

Porstendörfer 2004; Mohery et al. 2014) The short-lived progeny on the ambient particles then decay into two long-lived progeny, Lead-210 and Polonium-210, which respectively account for most of the beta- and alpha-radiation emitted by the particulate.(Cabello et al. 2018; Hernández et al. 2005) UOGD could influence local PR level by increasing the emission rate of radon. There is an increasing interest in the health effects of PR because the particle-bound Lead-210 and Polonium-210 tend to be deposited on the bronchial epithelium and expose neighboring cells to high-energy alpha particles that induce the carcinogenesis process.(Darby et al. 2005; Duan et al. 2015) Short-term exposure to PR has been associated with adverse health outcomes, including a decrease in lung function,(Nyhan et al. 2019) an increase in blood pressure,(Nyhan et al. 2018), and increased levels in biomarkers of inflammation.(Li et al. 2018; A. Blomberg et al. 2020)

To our best knowledge, this is the first study to estimate the impact of UOGD on PR. Considering the potential emissions from UOGD-associated activities, we hypothesize that there is an association between upwind UOGD activities and downwind PR. For hypothesis testing, we first calculate daily upwind UOGD well count, then investigate its relationship with the PR measured at a downwind monitor. Furthermore, we evaluate the distance-dependent decay of the influence. Finally, we estimate the different impacts of UOGD and COGD. The results of our study contribute to the currently limited knowledge regarding the influence of UOGD on PR.

Methods

Ambient Particle Radioactivity Measurements

We obtained PR data from the RadNet monitoring network, which is operated by the U.S. Environmental Protection Agency (EPA). This network measures the background environmental radiation levels in the air, precipitation, and drinking water under both routine and emergency

conditions. During the study period from 2001 to 2017, 157 RadNet sites (Figure 2-12) reported gross-beta measurements of various time lengths. Most RadNet monitors are located in metropolitan areas for better population coverage. At each site, total suspended particles (TSP) are collected using a high-volume sampler with a 4-inch diameter polyester fiber filter. Samplers are operated continuously for a 3- or 4-day integration. Filters are then sent to the National Analytical Radiation Environmental Laboratory (NAREL) for the measurement of gross-beta radiation.(Fraass 2015; U.S. Environmental Protection Agency (EPA) 2012) To create quasi-daily values from the 3- or 4-day integrated samples, we assigned the same beta-radiation level to each day of the collection period.

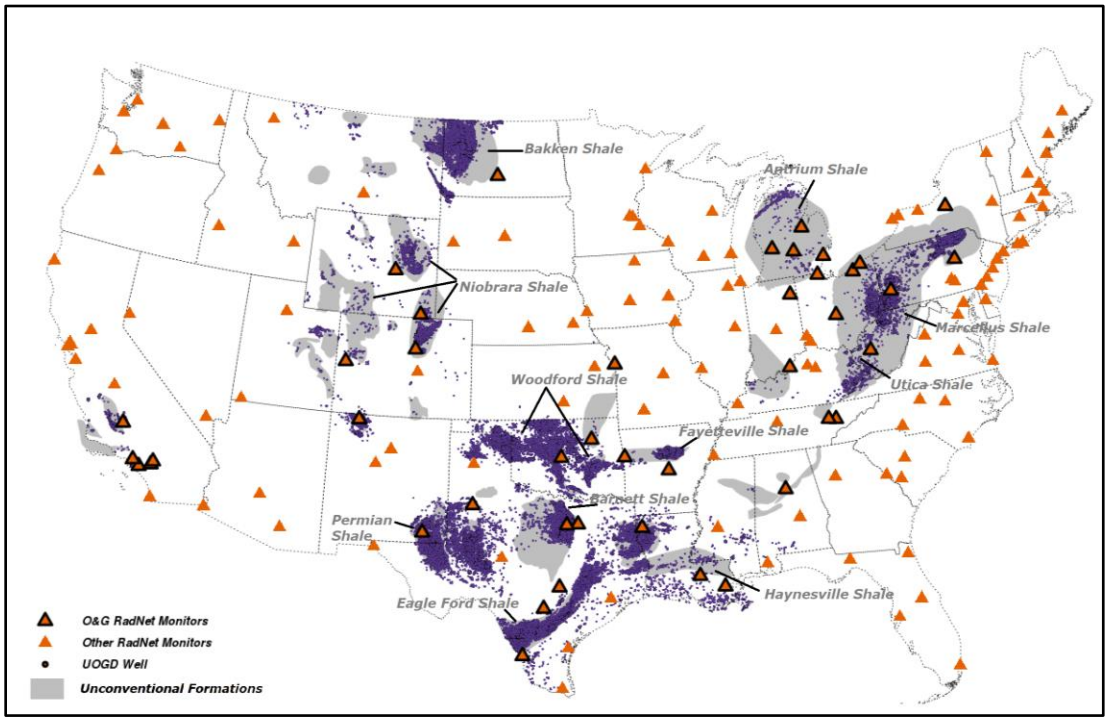


Figure 2-12. The location of RadNet monitors and the UOGD wells (completed by 2017) in the continental U.S

The 157 RadNet monitors are categorized into O&G RadNet monitor (black edge) and Other RadNet monitors (no edge) based on whether a monitor is within 50 km of any O&G extractive activities.

Unconventional Oil and Gas Development Data

We obtained position and production information of O&G wells from Enverus (formerly Drillinginfo), a third-party data vendor used by the Energy Information Administration (EIA) to prepare monthly fossil fuel production and marketing reports. The comprehensive data coverage of Enverus is achieved by compiling the permits, construction logs, and production records from state agencies. Details about this data source were presented in a previous study.(Czolowski et al. 2017) Our dataset includes information for 2,159,858 wells stimulated from 01/01/1949 to 12/31/2017. We used drilling type information as the primary indicator of whether a well is targeting an unconventional accumulation formation or not.(U.S. Enverionmental Preotection Agency (EPA) 2016) Specifically, we considered horizontally-drilled wells as UOGD wells and vertically-drilled wells as COGD wells. Directionally-drilled wells and wells without drilling type information were classified into UOGD or COGD wells based on their proximity to nearby UOGD wells and other secondary information using a Random Forest model (Supplementary Information **Section 1.1**).

We used the number of completed wells to characterize the intensity of O&G production activity. We identified a well as completed when the operator received the well from the driller. If this information is unavailable, we used the first production date as a proxy. Considering the transport of airborne particles, we focused on the completed wells positioned upwind of the RadNet monitor. Specifically, we created a circular sectional buffer centered on the daily wind direction with an angle of 90° and a radius of 20 km (Figure 2-13). We counted the numbers of UOGD and COGD wells, respectively, within the buffer on a daily basis to detect the different impacts of the two types of wells. We also created a series of circular sectional buffers at distances ranging from 25 to 50 km by 5 km intervals to evaluate the potential dependency of

impact on the spatial scale. For security reasons, the exact location of RadNet monitors is not publicly accessible. We used the centroid of each RadNet city as a proxy to the sampling location (Figure 2-13). Because of the potential spatial mismatch, we do not run any analysis on a spatial scale smaller than 20 km.

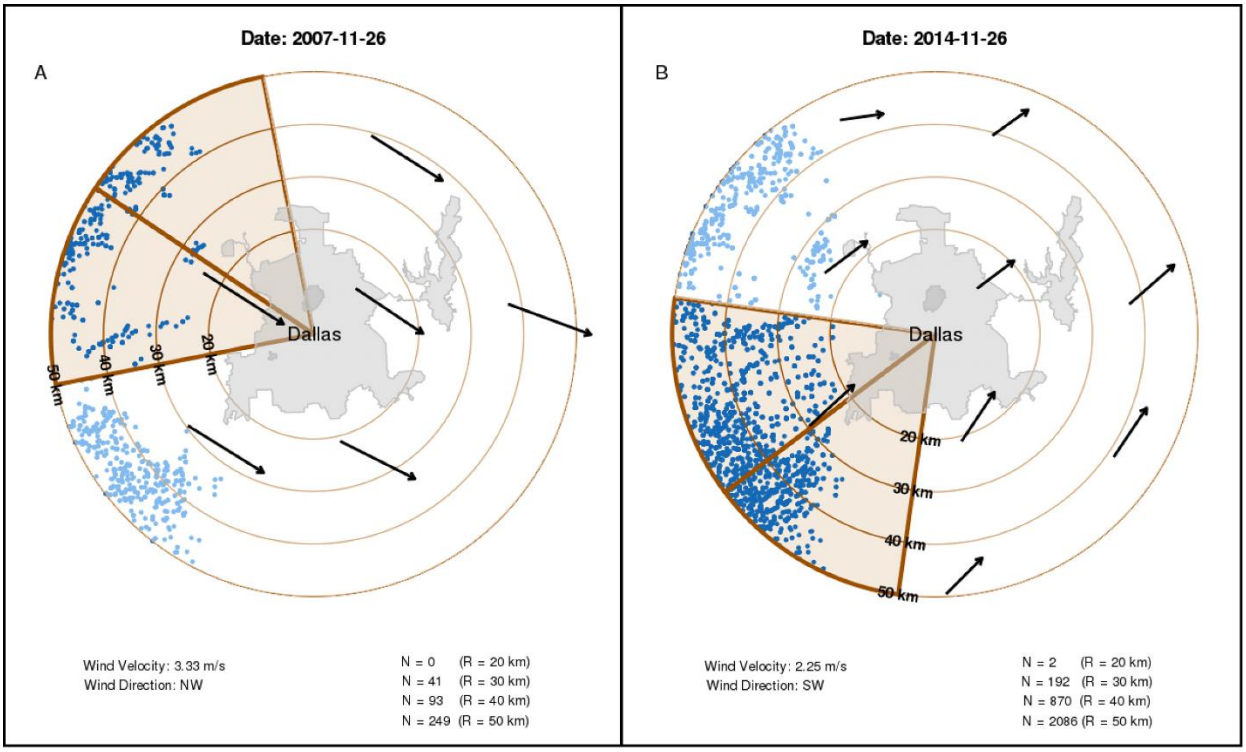


Figure 2-13. Methods to calculate the number of UOGD wells positioned upwind of the RadNet monitor at Dallas, TX in two example days

(**Panel A,** Nov-26-2007; **Panel B,** Nov-26-2014).

Due to the inaccessibility of the monitor's exact location, we used the geometric center of the city of Dallas, TX as a proxy. Based on daily wind direction (black arrows), we created the circular sectional buffer with a radius of 20 km and angle of 90 degree. We created a series of buffers with radiuses ranging from 20 km to 50km, in order to investigate the scale dependency. We used the same method to count the daily number of COGD wells upwind of RadNet monitors.

Predictors of Particle Radioactivity

We collected data on environmental variables related to the emission of PR. To control for the

emanation rate of radon from soil, we downloaded ground surface concentration of U-238 at a spatial resolution of 3 km from the United States Geological Survey.(Joseph S. Duval, John M. Carson, Peter B. Holman 2005) PR is associated with the origin of air masses because the emanation rate of radon from the ocean is two orders lower than that from the continent.(Baskaran 2011) To capture this pattern, we modeled four 72-hour back-trajectories (arrival time 06:00, 12:00, 18:00, and 24:00) of each RadNet monitoring sites using the Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT) model.(Stein et al. 2015) The proportion of trajectory over the continent was used as a proxy to the origin of the air mass with 0% indicating a maritime air mass, while 100% indicating a continental air mass. Finally, we collected the monthly number of sunspots observed by the Royal Observatory of Belgium,(Royal Observatory of Belgium 2019) as an indicator of the strength of solar activity. This information was used to adjust for the contribution of beta-emitting cosmogenic radionuclides originating from upper atmosphere.

We also obtained environmental factors influencing the transport of PR. Due to the scavenger effect of aerosol on the short-lived progeny of radon, $PM_{2.5}$ concentration strongly influences the spatiotemporal distribution of radioactive isotopes in the atmosphere.(Baskaran 2011) We downloaded daily $PM_{2.5}$ concentrations measured at EPA air quality monitors located within 50 km of RadNet sites and calculated the daily average if more than one measurement was carried out. The behavior of PR is driven by multiple meteorological factors, including wind velocity, relative humidity, planetary boundary layer height (PBLH), temperature, and soil moisture.(Hernández et al. 2005; Baskaran 2011) We obtained these variables from the North America Regional Reanalysis (NARR) dataset, with a spatial resolution of 32 km.(Mesinger et al. 2006)

Statistical Analysis

We applied linear mixed effect (LME) models to investigate the association between PR and UOGD activities. Dependency between daily PR measurements from the same monitor is controlled for by including monitor-specific random intercepts. We controlled for relevant environmental factors as fixed effects. We also adjusted for long-term temporal trend and seasonality by including polynomial terms based on the calendar year and temperature. We applied the LME as follows:

$\mathrm{P}R_{i,t} = (c_0 + \gamma_i)$	$+c_1 \cdot N\mathrm{um}_{i,t}$	
	$+c_2 \cdot U_i + c_3 \cdot origin_{i,t} + \mathrm{c}_4 \cdot sun_t$	: Emission of PR
	$+c_5 \cdot pm_{i,t} + c_6 \cdot pblh_{i,t}^{-1} + c_7 \cdot \mathrm{rhum}_{i,t}$ $+c_8 \cdot soilm_{i,t} + c_9 \cdot vel_{i,t}$	: Movement-related Environmental factors
	$+c_{10} \cdot lat_i$ $+\sum_{p=11,12} c_p \cdot year_t^{p-10} + \sum_{p=13\cdots15} c_p \cdot temp_{i,t}^{p-12}$	: Spatial and temporal trend
	$+c_{16} \cdot sunspots_t \times lat_i + c_{17} \cdot soilm_{i,t} \times U_i$ $+c_{18} \cdot Num_{i,t} \times vel_{i,t}$	: Interactions

where $\boldsymbol{PR_{i,t}}$ is PR level of site ***i*** on day ***t***; ***Num*** represents the number of upwind wells within the circular sectional buffer; $\boldsymbol{c_2}$ to $\boldsymbol{c_4}$ are the coefficients for the emission-dependent variables, including ***U*** for the ground surface concentration of U-238, ***origin*** for the origin of the air mass and ***sun*** for the number of sunspots; $\boldsymbol{c_5}$ to $\boldsymbol{c_9}$ are the coefficients for transport-related environmental factors, including ***pm*** for the concentration of $PM_{2.5,}$ $\boldsymbol{pblh^{-1}}$ for the inverse of HPBL, ***rhum*** for relative humidity, ***soilm*** for the moisture of soil and ***vel*** for wind velocity; $\boldsymbol{c_{10}}$ is

the coefficient for latitude-dependent spatial trend; $\boldsymbol{c}_{11}$ to $\boldsymbol{c}_{15}$ are the coefficients for temporal trends terms, represented by the polynomial terms of the calendar year (***year***) and temperature (***temp***); $\boldsymbol{c}_{16}$ to $\boldsymbol{c}_{18}$ are the coefficients for interactions terms between environmental factors.

In the primary analysis, we associated daily PR levels with the daily number of upwind UOGD wells within 20 km with LME. To investigate the magnitude of UOGD's impact, we calculated the increase in PR associated with the 95% percentile of upwind UOGD well count. To investigate the dependency of effects on the transport distance, we counted the number of O&G wells within a series of circular sectional buffers at distances ranging from 25 to 50 km by 5 km intervals, and then estimated the effects for each buffer distances.

To investigate the influence of COGD wells, we associated daily PR with the upwind number of COGD wells within the same buffers. As a negative control, we counted the number of downwind UOGD wells within 20 km. To identify the potential regional heterogeneity in the effects of UOGD, we restricted our analysis to three separate subregions named after the shale formations underneath: Marcellus-Utica subregions, Permian-Haynesville subregion, and Bakken-Niobrara subregion (Supplementary Information **Figure S3**). For sensitivity analysis, we first re-evaluated the associations by calculating the upwind UOGD wells within circular sectional buffers using two another central angles (60° and 120°). We then performed a leave-one-out sensitivity analysis to assess whether our estimates were sensitive to the omission of any single RadNet site.

We used the LME methods implemented in lme4 package (version 1.1-21)(Bates et al. 2015) in R (version 3.4.2)(R Core Team 2017) to fit the models. The significance test was based on confidence intervals instead of p-values. The analysis was conducted on the Cannon cluster,

supported by the Research Computing Group at Harvard University, Faculty of Arts and Sciences.

Results

Table 2-4. Descriptive statistics of the ambient particle radioactivity and its environmental predictors.

Variable	Nationwide	O&G Monitors	Other Monitors
Monitors (n)	157	43	114
Observations (n)	320,796	106,057	259,090
PR (mBq/m^3)	0.35 (0.22, 0.43)	0.39 (0.26,0.47)	0.33 (0.20,0.41)
^{238}U (ppm)	1.82 (1.58, 2.17)	1.91 (1.62,2.19)	1.74 (1.33,2.12)
Origin of Air Mass (%)	0.77 (0.57, 1.00)	0.82 (0.75,1.00)	0.75 (0.56,1.00)
$PM_{2.5}$ ($\mu g/m^3$)	9.93 (5.74, 12.70)	10.80 (6.49,13.50)	9.63(5.47,12.05)
Soil moisture (ton/m^2)	0.51 (0.42, 0.62)	0.52 (0.42,0.61)	0.52 (0.43, 0.62)
Relative humidity (%)	69.20 (62.1, 82.30)	65.80 (55.4,80.00)	71.30 (64.7,83.40)
Temperature (°C)	13.90 (6.70, 22.10)	13.90 (6.48,22.32)	13.48 (6.21,21.63)
PBLH (km)	0.91 (0.59, 1.13)	0.98 (0.66,1.18)	0.89 (0.57,1.11)
Wind speed (m/s)	3.36 (2.07, 4.39)	3.51 (2.14, 4.51)	3.27 (1.96,4.23)
Sunspots (n)	66.76 (19.00, 97.30)	--	--

Continuous environmental factors are summarized as the mean and the interquartile range (from 25th percentile to 75th percentile).

The unit of the gross-beta radiation is millibecquerel (mBq). 1 mBq/m3 is equal to 1×10-3Bq/m3. 1 mBq/m3 is also equal to 0.027 pCi/L

In this study, we analyzed 320,796 PR measurements carried out at 157 RadNet monitors across the continental United States from 2001 to 2017. We categorized these monitors into O&G RadNet monitors and other RadNet monitors, based on the existence of O&G extraction

activities within 50 km (Figure 2-12). As summarized in Table 2-4, the national average PR level is 0.35 mBq/m^3, with an interquartile range (IQR) from 0.22 mBq/m^3 to 0.43 mBq/m^3. O&G RadNet monitors have a higher average PR (0.39 mBq/m^3; IQR: 0.26,0.47 mBq/m^3), compared to the average PR of other O&G RadNet monitors (0.33 mBq/m^3; IQR: 0.20,0.41 mBq/m^3). Concerning PR emission-related environmental factors, O&G RadNet monitors have a higher ground surface U-238 level, and a higher percentage of air mass originated from the continent. For PR movement-dependent factors, O&G RadNet monitors have higher $PM_{2.5}$ concentrations, higher PBLHs, higher wind velocity, and lower relative humidity.

After excluding wells without production records, there are 1,574,602 completed O&G wells by the end of 2017. Out of these O&G wells, 152,904 (9.7%) are UOGD wells, and 1,421,698 (91.3%) are COGD wells (Supplementary Information Figure 2-S16). Out of the UOGD wells, 4,611 (3.0%) are within 20 km of RadNet monitors and 28,016 (18.3%) are within the 50 km buffer. UOGD expanded rapidly in all three subregions during the study period (Supplementary Information **Section 1.3**). Fort Worth, Texas, had the highest average upwind UOGD count (mean 586, IQR: 504,661) within 20 km in 2017.

According to our LME model, there is a statistically significant association between the downwind PR and the upwind UOGD activity. With adjustment for environmental factors regarding the natural emission and movement of PR, an additional 100 upwind UOGD wells within 20 km is associated with a 0.024 mBq/m^3 increase in the level of PR (95% CI: 0.020, 0.028 mBq/m^3) for a wind velocity of 1m/s (Supplementary Information **Section 2.1**). Under the same wind condition, an additional 100 upwind COGD wells within 20 km is associated with a 0.004 mBq/m^3 increase in PR (95% CI: 0.003, 0.004 mB1/m^3). Regarding the magnitude of the impact, UOGD and COGD could elevate the PR level by up to 0.13 mBq/m^3 and 0.029 mBq/m^3,

respectively. We also found significant negative interactions between wind velocity and the counts of both UOGD wells and COGD wells located upwind of RadNet monitors, suggesting lower influence when the wind is strong (Table 2-5).

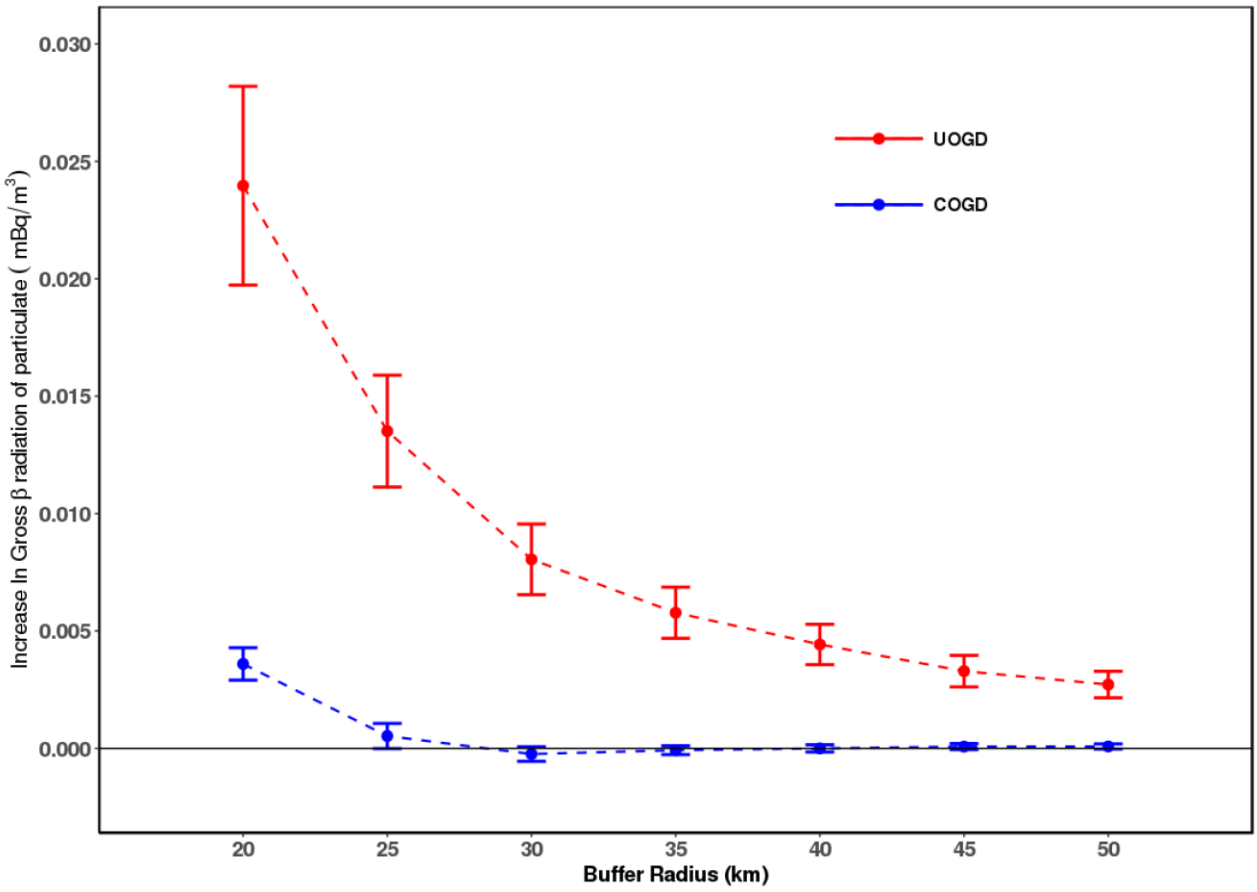


Figure 2-14.The association between upwind O&G production activities and downwind level of PR.

The increase in PR associated with an additional 100 UOGD wells (blue bars) and COGD wells (red bars) at multiple buffer distance. Effect estimations are visualized as the points and the 95% Cis are visualized as the bars. The source data of this figure is provided in Supplementary Materials **Table S2**.

Table 2-5**.** The associations between PR and other environmental factors

Term	Estimation	95% CI	Details
$U\ (10^{-2})$	4.88	(3.29, 6.48)	U-238 level in ground surface material
$origin\ (10^{-2})$	7.13	(6.85, 7.41)	The origin of the air mass. 1 indicates purely continental air mass; 0 indicates purely oceanic air mass
$temp\ (10^{-1})$	-3.73	(-4.48, -2.99)	The polynomial terms of air temperature. These are used to adjust for seasonality.
$temp^2(10^{-3})$	1.20	(0.93, 1.46)	
$temp^3(10^{-6})$	-1.27	(-1.58, -0.96)	
$year(10^{-3})$	-1.63	(-2.21, -1.05)	The polynomial terms of the calendar year. These are used to control for long-term trend
$year^2(10^{-6})$	9.29	(-18.72, 37.31)	
$PBLH^{-1}$	16.53	(15.50, 17.64)	Inverse of PBLH
$pm(10^{-2})$	1.12	(1.11, 1.14)	Average concentration of $PM_{2.5}$ within 50 km from the RadNet monitor.
$rhum(10^{-4})$	-3.04	(-3.50, -2.58)	The relative humidity 2 meters above surface
$soilm(10^{-5})$	-4.84	(-6.46, -3.21)	Liquid volumetric soil moisture in the top 1m of soil
$sun(10^{-4})$	1.45	(1.23, 1.67)	Monthly number of sunspots
$lat(10^{-3})$	-6.84	(-9.51, -4.18)	The latitude of the RadNet monitor.
$vel(10^{-3})$	-0.91	(-1.37, -0.45)	Wind velocity 10 m above surface.
$U{\cdot}soilm(10^{-5})$	-6.57	(-7.43, -5.72)	The interaction between U-238 concentration and soil moisture
$sun{\cdot}Lat(10^{-5})$	-2.52	(-2.90, -2.15)	The interaction between monthly count of sunspots and latitude

Furthermore, we found that PR is significantly associated with upwind UOGD at every buffer distance of our study (Figure 2-14). However, the influence of an additional 100 UOGD wells decreased gradually as the buffer radius increased from 20 km to 50 km. An additional 100 UOGD wells within 50 km is associated with a 0.002 mBq/m^3 increase in PR (95% CI: 0.002, 0.003 mBq/m^3). Meanwhile, the association between the upwind COGD well count and PR is

not statistically significant when the buffer radius is greater than 20 km.

In the negative control analysis, we found PR level is statistically associated with the number of UOGD wells within 20 km downwind of a RadNet monitor. However, the increase of PR associated with an additional 100 downwind UOGD wells within 20 km is 0.021 mBq/m^3 (95% CI: 0.017, 0.024 mBq/m^3), which is smaller than the impact of an additional 100 upwind UOGD well (0.024 mBq/m^3 ,95% CI: 0.020, 0.028 mBq/m^3). We found that our results are not sensitive to the angle of the buffers (Supplementary Information **Section 2.2**, Figure 2-S18). The upwind-downwind difference in the influence of UOGD wells is more significant when the angle of the circular sections is 60° compared to the quadrant buffer used in the primary analysis (Supplementary Information Table 2-S9). We also found that our results are robust to omitting one

of the RadNet monitors in the analysis (Supplementary Information **Section 2.3,** Figure 2-S19).

We found that the PR level is statistically associated with the emission- and movement-dependent environmental covariates (Table 2-5). There are significant positive correlations between PR and the ground surface concentration of U-238, the proportion of continent-sourced air mass, the number of sunspots, $PM_{2.5}$ concentration, and the inverse of PBLH. Meanwhile, PR is negatively associated with latitude, relative humidity, and soil moisture.

In our subregional analysis, we found significant heterogeneity among the three subregions regarding the influence of UOGD on PR (Table 2-6). In the Marcellus-Utica subregion, PR is not statistically significantly associated with upwind UOGD wells for any buffer distances investigated. However, the association is significant for each buffer distance in the Permian-Haynesville subregion. In the Bakken-Niobrara subregion, the association is not statistically

significant when the buffer radius is smaller than 30 km.

Table 2-6**.** The associations between PR and upwind UOGD well count in three subregions of our study extent.

	Marcellus-Utica Subregion		Permian-Haynesville Subregion		Bakken-Niobrara Subregion		The whole study extent	
Radius (Km)	Est (10^{-2})	95%CI (10^{-2})	Est (10^{-2})	95%CI (10^{-2})	Est (10^{-2})	95%CI (10^{-2})	Est (10^{-2})	95%CI (10^{-2})
20	17.96	(-3.06, 38.99)	1.33	(4.85,7.80)	1.26	(-4.52,13.79)	2.40	(1.97, 2.82)
25	2.05	(-2.95, 7.05)	0.81	(2.79,4.12)	0.93	(-3.24,11.07)	1.35	(1.11, 1.59)
30	0.24	(-1.16, 1.63)	0.54	(1.81,2.87)	3.61	(3.87,17.32)	0.81	(0.65, 0.96)
35	0.11	(-0.57, 0.79)	0.38	(1.20,1.89)	3.64	(5.72,12.93)	0.58	(0.47, 0.69)
40	0.08	(-0.35, 0.51)	0.29	(0.91,1.42)	2.12	(3.42,7.81)	0.44	(0.36, 0.53)
45	0.07	(-0.24, 0.38)	0.21	(0.65,1.06)	1.51	(2.53,5.80)	0.33	(0.26, 0.40)
50	0.09	(-0.17, 0.34)	0.17	(0.50,0.86)	1.05	(1.81,3.71)	0.27	(0.22, 0.33)

The estimated influence (shown in columns named as Est) is presented as the increase in PR associated with an additional 100 UOGD wells within the radius.

Discussion and Conclusion

In this study, we analyzed the radioactivity of airborne particles collected at 157 RadNet monitors across the continental United States from 2001 to 2017 (Figure 2-12). To characterize upwind UOGD activities, we used the position and production records of 152,904 UOGD wells and counted the daily number of upwind UOGD wells. Our results add to the limited literature by evaluating the influence of UOGD on the radioactivity of ambient particles.

These associations suggest the existence of some pathways by which UOGD activities could release NORM into the atmospheric environment. Likely mechanisms include the fugitive release of natural gas, which contains a higher-than-background level of radon at wellheads,

compressor stations, pipelines, and other associated facilities;[48–50] the management, storage, discharge and disposal of flow-back and produced water which is rich in NORMs;[16,51–53] the accidental spill or beneficial use of produced water in nearby communities;[22] the handling, transport, management, and disposal of radioactive drill cuttings.[18,19] To distinguish the contributions of these surface activities, more continuous measurements of PR, especially for some specific radionuclides, are needed at a finer spatiotemporal resolution.

Our results show a remarkable distinction between the impacts of UOGD and COGD on PR. UOGD-specific processes, such as hydraulic fracturing and directional drilling, could potentially explain the larger associated impacts. The hydraulic fracturing process produced large volumes of flow-back water and drilling mud, which are subsequently stored in the temporary reserve pit adjacent to the drilling site. Most UOGD production states allow the operator to close the reserve pit within up to one year after completing the drilling.[54] This practice potentially enables the NORMs in the produced water to decay into radon above the ground surface and release the radon into the ambient environment. The lateral drilling process produces large volumes of drill cuttings from the unconventional accumulating formation, whose levels of NORMs are higher than those produced during the vertical drilling stage. These drill cuttings are currently not considered hazardous wastes by U.S. EPA. The practice of beneficial use of drill cuttings and land treatment could potentially release radon into the ambient environment.[55]

Our subregional analysis demonstrates remarkable heterogeneity in the estimated influences of UOGD in the three subregions. Due to a lack of monitors with UOGD wells nearby, the Marcellus-Utica subregional model did not have enough power to detect statistically significant associations. For a buffer radius of 20 km, the estimated influence of an additional 100 upwind UOGD wells is 17.96×10^{-2} mBq/m^3 (95% CI: -3.06, 38.99 $\times10^{-2}$ mBq/m^3) in the Marcellus-Utica

subregion, approximately seven times the estimated effects of a nationwide model. The difference is likely caused by the relatively few UOGD wells near the RadNet monitors in the Marcellus-Utica subregion (Supplementary Information **Section 1.3,** Table 2-S7). In the Bakken-Niobrara subregion, only two RadNet monitors (Casper, WY, and Navajo Lake, NM) have active UOGD wells around when the buffer radius is smaller than 30km. When we enlarged the radius to 30 km, two additional RadNet monitors (Denver, CO, and Grand Junction, CO) had UOGD within the buffer, enabling us to identify the significant association (Supplementary Information Figure 2-S17).

Our results show a monotonic declining impact of O&G wells on PR as the buffer radius increases from 20 to 50 km (Figure 2-14). The attenuation of radon after being emanated can explain this pattern. The trend indicates a more significant influence on the PR level of communities close to intensive UOGD activities. Limited by the accuracy of RadNet monitor location information, we did not estimate the impacts on a buffer distance smaller than 20 km. To tentatively extrapolate our results to these neighborhoods, we modeled the estimated influences as a power function of the radiuses with a negative exponent (Supplementary Information **Section 3.1,** Figure 2-S20). Based on this tentative extrapolation, an additional 100 UOGD wells within 10 km would be associated with an increase of 0.14 mBq/m^3. However, the result of this extrapolation should be interpreted cautiously. Monitors closer to UOGD wells are needed to validate this extrapolation.

One strength of our study is the nationwide monitor network of PR. The long-term measurement enables us to compare the current PR level with the baseline PR level in the absence of widespread UOGD. Furthermore, all filters were measured by NAREL using the same protocol,

excluding the uncertainties induced by the heterogeneous devices operated by different labs. The other strength is the comprehensive database covering O&G activities. The Enverus database facilitated distinguishing the distinct impacts of UOGD and COGD on PR. In addition, we obtained diverse environmental covariates related to the natural emissions and transport of PR. Adjustment for these factors allowed us to draw conclusions explicitly related to the impacts of O&G development by explaining the natural variation of PR. The associations between PR and these environmental factors, as summarized in Table 2-5, are in agreement with the findings of previous studies.[26,56]

One limitation of this study is that we only associated PR with the existence of completed O&G wells. Other construction-dependent factors may also influence the emission rate. However, the O&G wells with a detailed construction records are rare in our database, making it difficult to know the duration of construction, thus limiting our ability to investigate the construction-dependent association. Another limitation of our study is the simplification of the particle transport process. Our circular-sectional buffers are designed based on the Gaussian Dispersion Model. This calculation assumed a steady-state meteorological condition, which is reasonable in this case due to the short downwind distances. However, this computation could be improved by introducing advanced atmospheric dispersion models.

Our results indicate the significant influence of UOGD on PR, a previously overlooked property of $PM_{2.5}$. Particulate-bound radon progeny continue releasing ionizing radiation after being inhaled and thus could induce systemic oxidative stress and inflammation, even at the levels observed in this study. Nyhan et al. (2018 and 2019) found that a 0.07 mBq/m^3 increase in 28-day average gross-beta radiation is associated with a 2.95 mm-Hg increase in diastolic blood pressure, a 3.94 mm-Hg increase in systolic blood pressure, a 2.41% decrease in forced vital

capacity, and a 2.41% decrease in forced expiratory volume in Normative Ageing Study (NAS) population.[34,35] Blomberg et al.(2020) reported that a 0.12 mBq/m^3 increase in seven-day average gross-beta radiation is associated with a 4.9% increase in C-reactive protein, a 2.8% increase in intercellular adhesion molecule-1, and a 4.3% increase in vascular cell adhesion molecule-1 in the same study popultaion.[37] Jointly with these associations, our results indicate that increase in PR due to the extensive UOGD could have already caused adverse health outcomes in nearby communities by elevating PR level (Supplementary Information **Section 3.2**).

Conclusion

Our analysis demonstrates that upwind UOGD activities could significantly elevate the PR level in downwind communities. UOGD has a larger impact on PR, compared to COGD. Based on previously published health effect analysis of PR, it is possible that the widespread of UOGD could induce adverse health effects to residents in proximity by elevating the PR.

Appendix

Methods and Materials

Prediction of Drilling Type

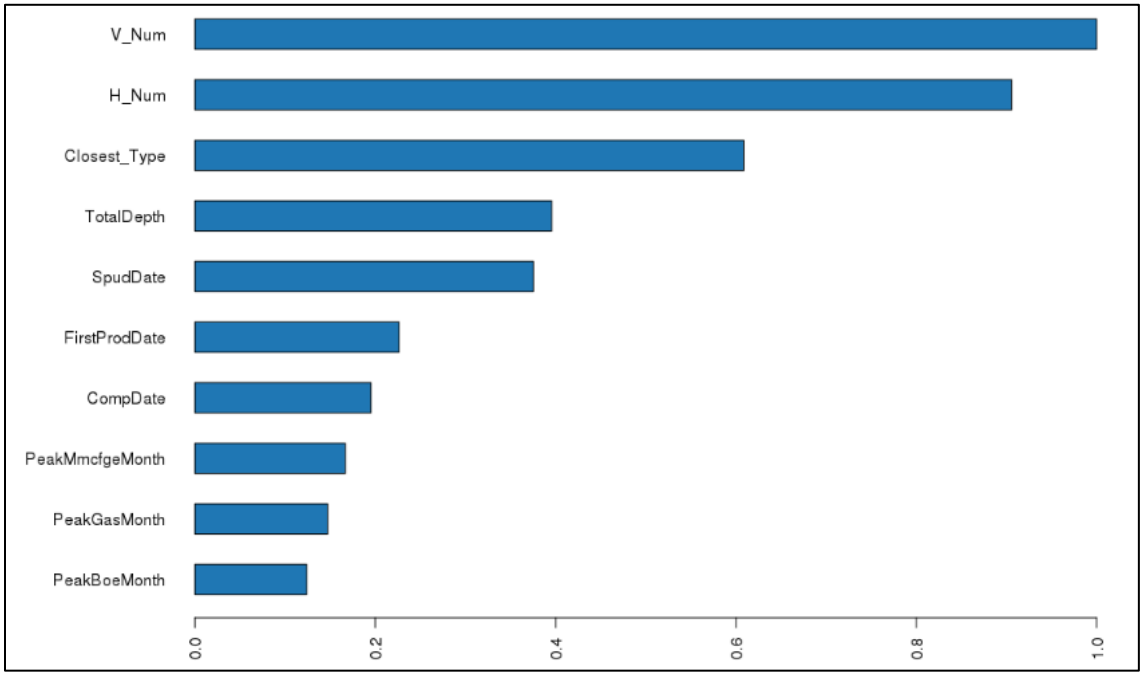


Figure 2-S15**.** Relative importance of covariates in the random forest model.

Full names of the variables from top to bottom: Number of conventional wells within 10km, number of unconventional wells within 10 km, the drilling type of the closest well with known drilling type, total drilling depth, spudding date, the date of the first production record, completion date, the date of peak million cubic feet of gas equivalent production, the date of peak natural gas production, the date of peak barrel of oil equivalent production

State energy agencies are the primary data source of Enverus (formerly Drillinginfo). Directional drilling has different definitions among states. For example, drilled wells and horizontally drilled wells are both grouped in a single class in Colorado as directional wells, while the two are separate in New Mexico. As a result, there is a visually remarkable difference in the percentage of horizontal wells, which are considered unconventional wells, across the state line even though they

share the target geological formation. As a result, it is not reliable to assume that all directionally drilled wells are unconventional wells nationally. Besides, the raw dataset from Enverus does not provide drilling type information for more than 75% of the wells, mostly drilled before 2000. However, it is also inaccurate to assume all wells without drilling type information are conventional because drilling type is not required by a state agency but reported by the operators voluntarily. Almost all wells in Alabama do not have drilling type information. To solve these problems, we need to predict the binary drilling type based on known drilling types of nearby wells and other secondary information.

We fitted a random forest model to perform this prediction. Random forest is a regression tree-based algorithm good at capturing the non-linear relationship between the primary variable and secondary variables, thus suitable for solving this binary classification problem[1]. Secondary variables incorporated in the model include: 1) drilling type of the nearest well with known drilling type information; 2) distance to the nearest conventional/unconventional well; 3) fractions of conventional and unconventional wells of the nearest 10 wells with known drilling type; 4) O&G reservoir where the well is positioned; 5) spudding/completion time; 6) drilling depth; 7) natural gas /liquid production in the first 6 months, and; 8) production declining rate of gas/liquid. After running a grid search for optimal performance, the parameters of this model were set as follows: the number of trees was 100, the maximum depth was 15, the minimum size of node was 5. The accuracy of this model was **99.83%** for COGD wells and **93.1%** for UOGD wells. The performance difference is potentially caused by the re-fracturing process of some conventional O&G wells. As shown in Figure 2-S15, the 10 most important covariates were: number of COGD wells within 10 km, number of UOGD wells within 10 km, the drilling type of the closest well with known drilling type, the total drilling depth (vertical drilling length plus lateral drilling

length), spudding date, the first production date, completion date, the date of peak million cubic feet of gas equivalent production, the date of peak natural gas production, the date of peak barrel of oil equivalent production.

We used the Distributed Random Forest (DRF) method implemented in h2o package (version 3.26.0.2) in R (version 3.4.2) to fit the models.

Location of COGD Wells

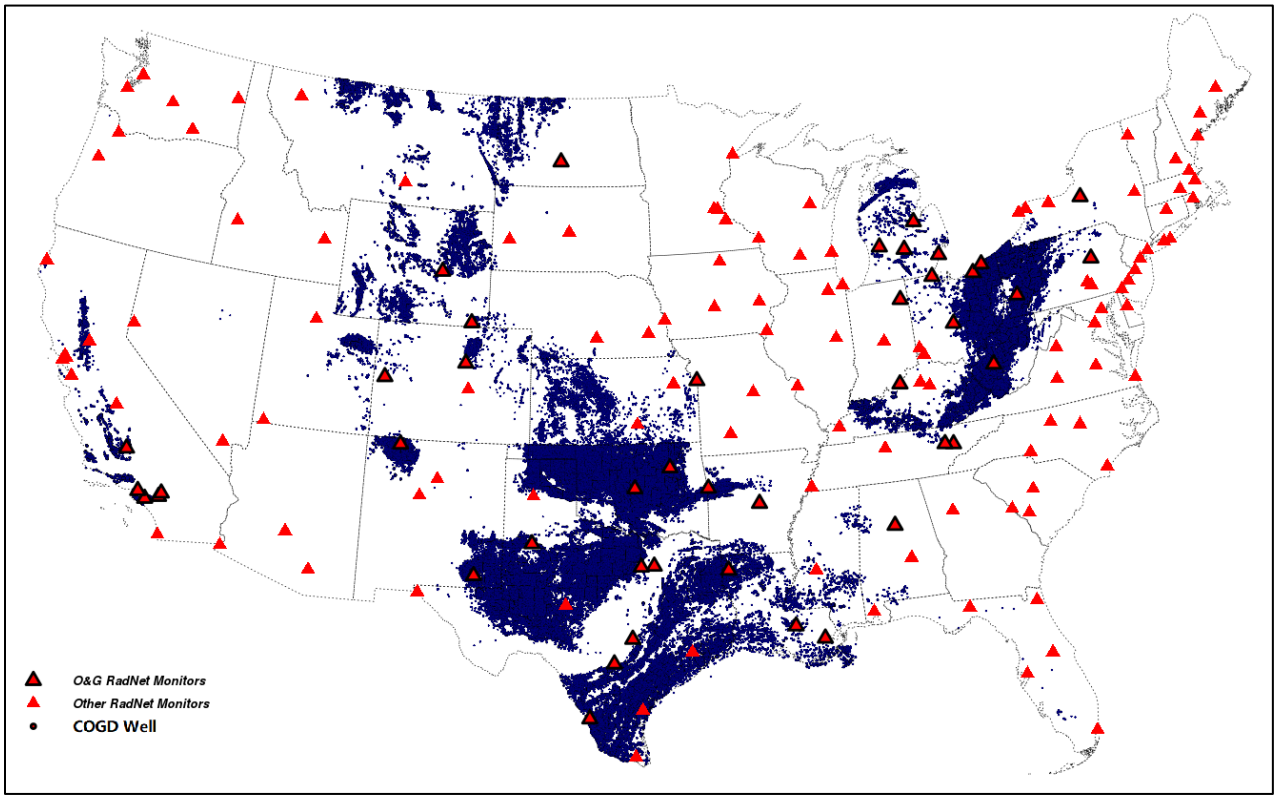


Figure 2-S16. The location of COGD wells completed by December 2017.

Temporal Trend of UOGD

Most O&G wells, especially UOGD wells, are relatively distant from metropolitan areas where the RadNet monitors are located. As a result, the average upwind UOGD well count does not fully represent the temporal trend of UOGD expansion. As summarized in Table 2-S7, we

selected one RadNet monitor from each of the three subregions to represent the temporal trend of upwind UOGD well count.

Table 2-S7**.** Annual average upwind UOGD count in three representative cities of three subregions.

	Pittsburg, PA		Fort Worth, TX		Navajo Lake, NM	
	20 km	50 km	20km	50km	20km	50km
2000	0.0	0.0	0.0	0.0	0.0	0.0
2001	0.0	0.0	0.0	0.0	0.0	0.0
2002	0.0	0.0	0.0	0.0	0.0	0.0
2003	0.0	0.0	0.0	0.0	0.0	0.0
2004	0.0	0.0	0.0	0.0	0.0	0.0
2005	0.0	0.0	0.0	0.0	0.0	0.0
2006	0.0	2.5	0.0	0.0	0.0	0.0
2007	0.0	5.7	143.1	980.2	0.0	0.0
2008	0.0	24.0	216.6	1561.0	0.0	0.0
2009	0.0	41.2	293.6	2020.6	0.0	0.0
2010	0.0	105.5	372.4	2502.8	43.0	111.5
2011	1.0	174.1	466.3	3076.5	46.9	112.6
2012	1.0	232.0	523.6	3179.9	47.6	135.0
2013	1.9	314.3	533.4	3060.2	53.8	155.8
2014	11.8	380.8	559.0	3143.0	45.0	141.3
2015	12.0	456.0	572.0	3196.5	55.1	154.9
2016	12.0	540.2	573.1	3095.0	56.9	155.7
2017	12.0	575.1	585.8	3240.9	59.3	162.8

Extent of Subregions

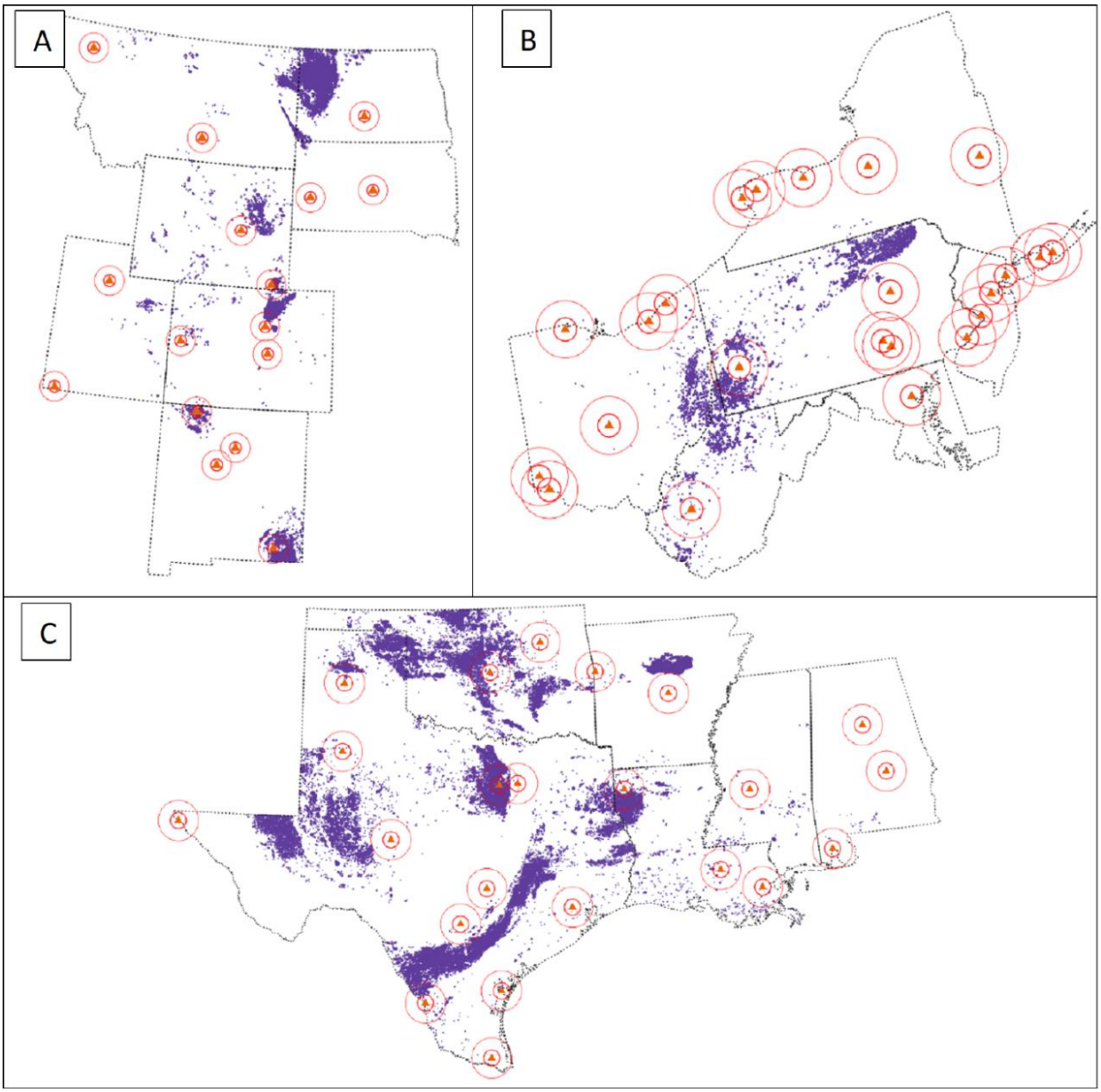


Figure 2-S17. Subregions of our study extent, primarily determined by the shale formations.

Inner circles represent the circular buffers with a radius of 20 km. Outer circles represent the circular buffers with a radius of 50 km. Panel A shows the extent of Bakken-Niobrara subregion, which includes MT, ND, SD, WY, UT, CO, and NM; Panel B shows the extent of Marcellus-Utica subregion, which includes PA, OH, WV, NY, NJ, and MD; Panel C shows the extent of Permian-Haynesville subregion which covers TX, OK, AR, LA, MS, and AL.

Results

Results of Primary Analysis

Table 2-S8**.** Source data for Figure 3 in the main text

Variable	Radius (km)	Coefficient (mBq/m^3 per 100 wells)	95% CI Lower Bound	95% CI Upper Bound
Upwind UOGD	20	2.40E-02	1.97E-02	2.82E-02
Upwind UOGD	25	1.35E-02	1.11E-02	1.59E-02
Upwind UOGD	30	8.05E-03	6.54E-03	9.55E-03
Upwind UOGD	35	5.78E-03	4.69E-03	6.87E-03
Upwind UOGD	40	4.43E-03	3.57E-03	5.28E-03
Upwind UOGD	45	3.29E-03	2.62E-03	3.96E-03
Upwind UOGD	50	2.72E-03	2.15E-03	3.28E-03
Upwind COGD	20	3.60E-03	2.91E-03	4.29E-03
Upwind COGD	25	5.27E-04	-9.41E-06	1.06E-03
Upwind COGD	30	-2.39E-04	-5.49E-04	7.02E-05
Upwind COGD	35	-7.67E-05	-2.64E-04	1.10E-04
Upwind COGD	40	9.85E-08	-1.52E-04	1.53E-04
Upwind COGD	45	7.63E-05	-5.22E-05	2.05E-04
Upwind COGD	50	8.15E-05	-2.77E-05	1.91E-04

Sensitivity to the Angle of Circular Sectional Buffer

In our primary analysis, we counted the number of O&G wells within circular sectional buffers, whose central angle was 90° (As shown in Figure 2-13). We re-calculated the upwind UOGD

well in buffers with two additional central angles (60° and 120°), to test the sensitivity of our estimated effects. As shown in Figure 2-S18, our results are not sensitive to the variation in angle. The estimated effects for a smaller angle (60°) were larger than our original results, while the estimated effects for a larger angle (120°) were smaller than our original results. The negative correlation between the estimated effects and buffer angle is in agreement with the atmospheric dispersion model.

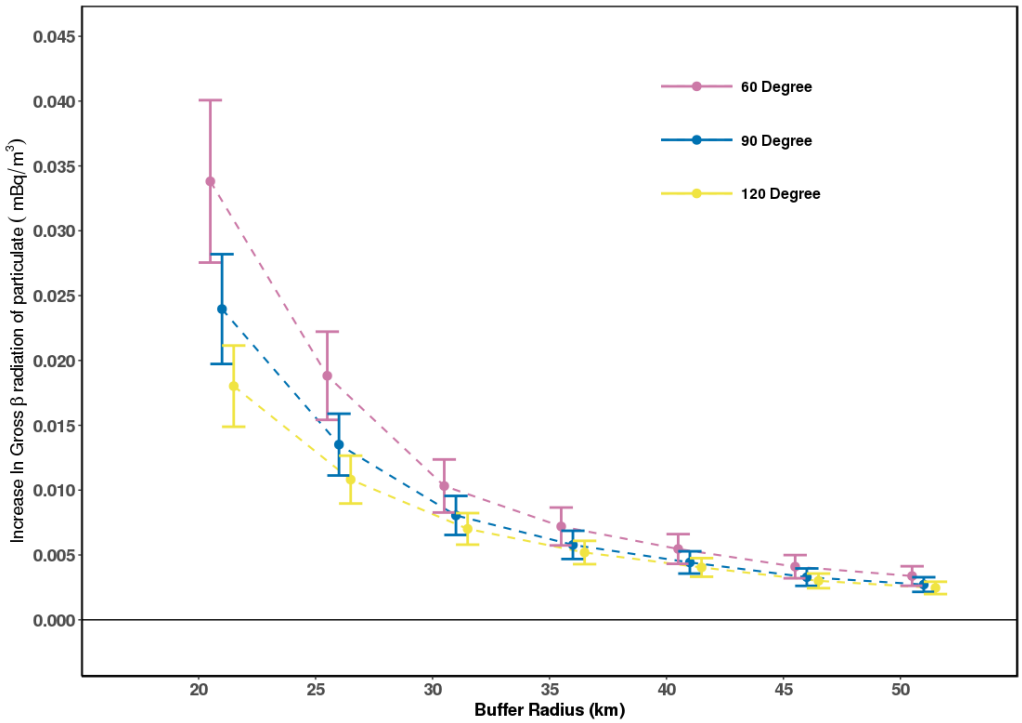


Figure 2-S18**.** The increment in PR associated with an increase of 100 upwind UOGD wells in circular sectional buffers with different central angles.

We visualized the estimated effects as points and 95% CI as bars. The source data is attached in Table 2-S9.

Table 2-S9**.** The source data table of Figure 2-S18.

Central Angle of the buffer (Degree)	Radius (km)	Coefficient (mBq/m^3 per 100 wells)	95% CI Lower Bound	95% CI Upper Bound
30	20	3.38E-02	2.76E-02	4.01E-02
45	20	2.40E-02	1.97E-02	2.82E-02
60	20	1.80E-02	1.49E-02	2.12E-02
45	25	1.35E-02	1.11E-02	1.59E-02
30	25	1.88E-02	1.54E-02	2.22E-02
60	25	1.08E-02	8.97E-03	1.27E-02
45	30	8.05E-03	6.54E-03	9.55E-03
30	30	1.03E-02	8.27E-03	1.24E-02
60	30	7.01E-03	5.80E-03	8.23E-03
45	35	5.78E-03	4.69E-03	6.87E-03
30	35	7.20E-03	5.74E-03	8.66E-03
60	35	5.19E-03	4.29E-03	6.09E-03
45	40	4.43E-03	3.57E-03	5.28E-03
30	40	5.46E-03	4.32E-03	6.61E-03
60	40	4.04E-03	3.32E-03	4.76E-03
45	45	3.29E-03	2.62E-03	3.96E-03
30	45	4.10E-03	3.20E-03	4.99E-03
60	45	3.00E-03	2.44E-03	3.57E-03
45	50	2.72E-03	2.15E-03	3.28E-03
30	50	3.38E-03	2.62E-03	4.13E-03
60	50	2.46E-03	1.99E-03	2.94E-03

Sensitivity to Omitting a Monitor

We carried out a leave-one-out analysis to investigate the likelihood that our estimated result is driven by a single RadNet monitor. Specifically, we iteratively exclude all PR measurements of a RadNet monitor and re-estimate the estimated effect with the remaining monitors using the same model formation. As shown in Figure 2-S19, our result was not sensitive to omitting a single RadNet monitor.

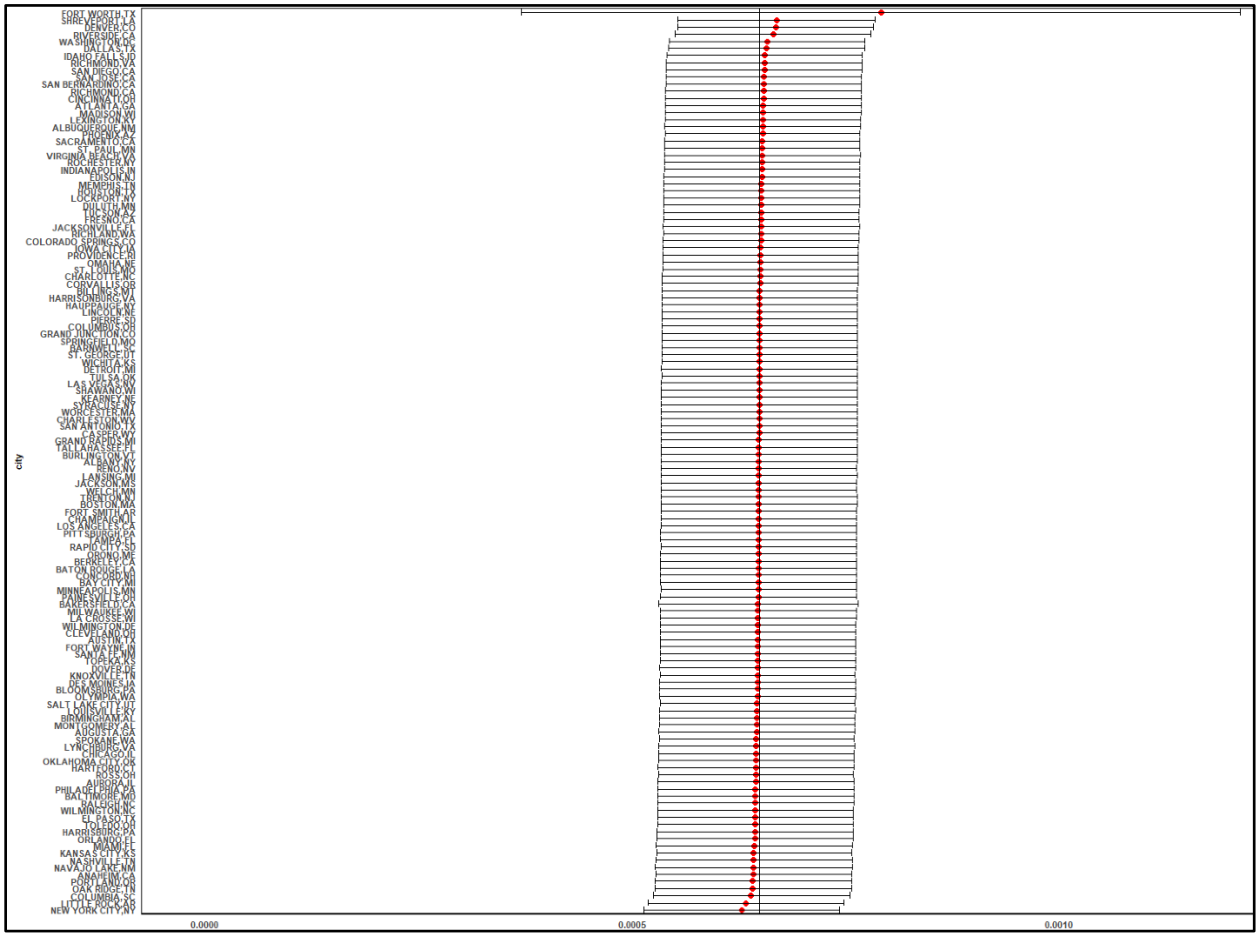


Figure 2-S19**.** The influence of omitting one monitor on the estimated association between PR and upwind UOGD well count within 20 km.

Discussion

1.1 Extrapolation of the Results

We observed an apparent distance-dependent decay in the effects of UOGD on PR (Figure 2-14). This trend suggests that we could tentatively extrapolate our results to a finer spatial scale. We used a power function with a negative exponent to fit the decay. As shown in Figure 2-S20, the exponent with the best fit is -2.5, between -2 (indicates a two-dimensional dispersion) and -3 (indicates a three-dimensional dispersion).

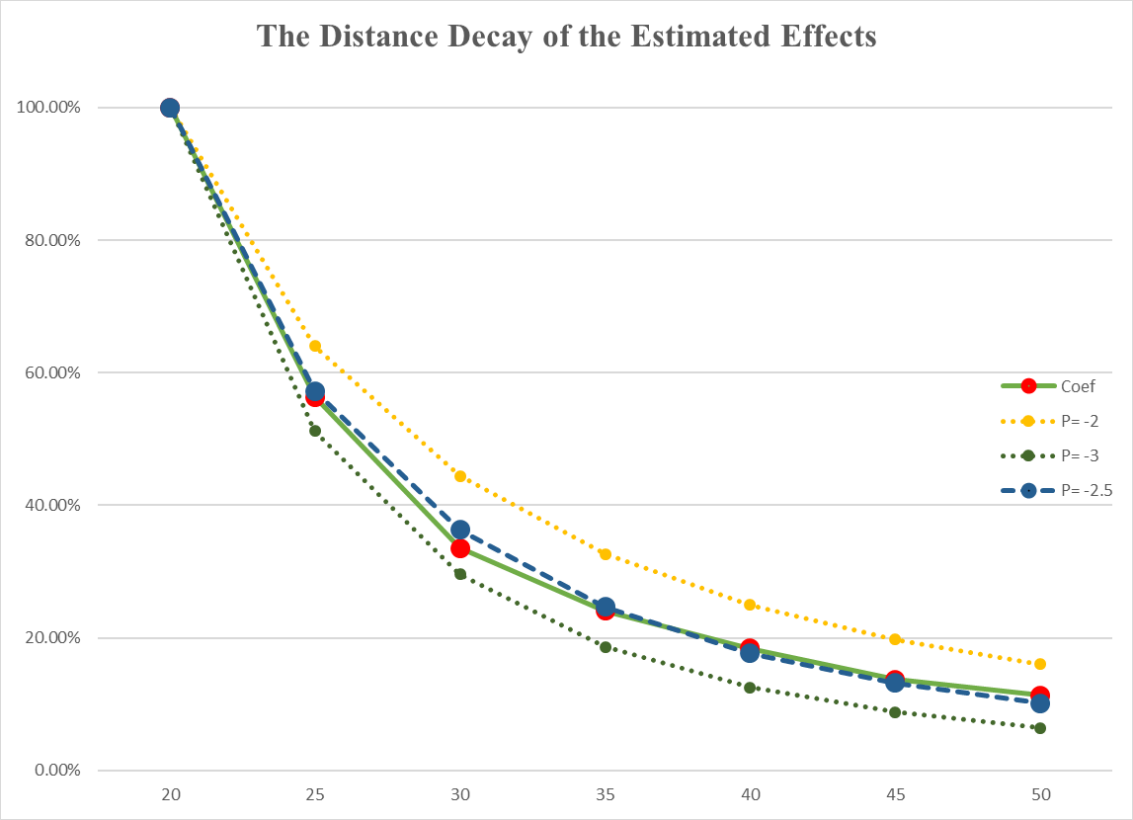


Figure 2-S20. The observed distance-dependent decay of the estimated effects and the modeled distance decay by power functions with negative exponents.

Y-axis represents the ratio between the estimated effects at the specific distance and the estimated effects at 20 km. The solid curve indicates the observed distance-decay of the estimated effects ("coef").

1.2 The Scale of the Influence

As shown in Figure 2-12, the majority of UOGD wells are not drilled close to metropolitan regions. Likely, the PR levels in some communities distant from RadNet monitors could also be elevated by the extensive UOGD activities nearby. To identify these communities, we mapped the 1 km grids whose annually-averaged upwind UOGD well count was over 580 (the 95% percentile of upwind UOGD well count from the modelling area) and over 100 at the end of 2017. As shown in Figure 2-S21, there are several regions, including the core part of Eagle Ford shale, Barnett shale, Fayetteville shale, and Niobrara shale, with an annual average UOGD over 580. The total area of these regions is 10,724 km^2. Over 1.2 million people reside in these areas. Based on the health studies cited in the main text, residents living in these communities are more likely to show adverse symptoms, including higher blood pressure, decreased lung function, and increased level of the inflammatory biomarker, if the PR increase downwind of UOGD wells in these regions are consistent with those of our study.

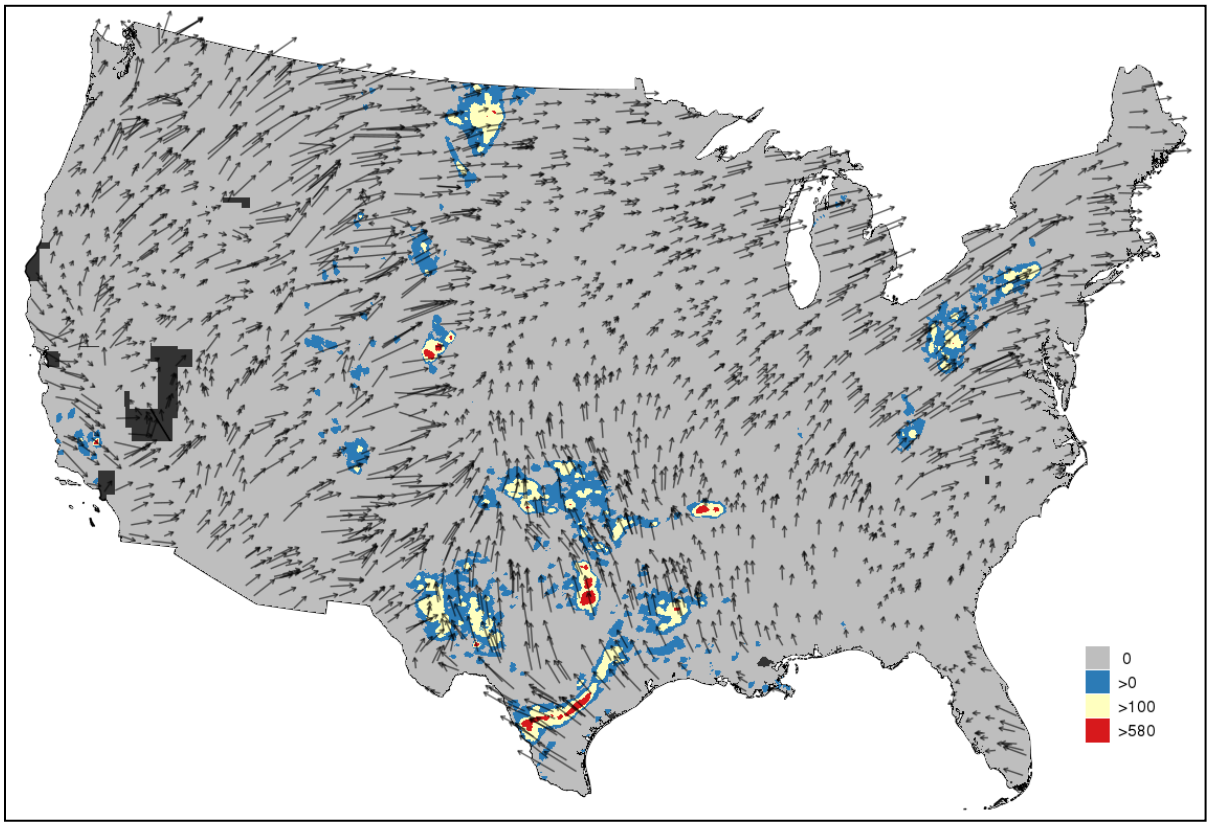


Figure 2-S21. Regions with annual average upwind UOGD wells over 580.

Wind vectors (black arrows in the figure) indicate the annual prevailing wind field, which was used to calculate upwind O&G development. UOGD wells used to calculate upwind activities were completed before 12/31/2017.

CHAPTER 3: Causal Associations of Residential Radon Exposure with All-cause Mortality in Medicare Beneficiaries in New England

Longxiang Li, M.S., Francesca Dominici, Ph.D., Annelise J. Blomberg, Sc.D., Joel D. Schwartz, Ph.D., Brent A. Coull, Ph.D., John D. Spengler, Ph.D., Yaguang Wei, M.S., Carolina L.Z. Vieira, Ph.D., Adjani A. Peralta, M.S., Petros Koutrakis, Ph.D.

Affiliations:

From the Department of Environmental Health (L.L, A.J.B, John D. Spengler, Joel D. Schwartz, B.A.C, W.Y, C.L.Z.V, A.P., P.K) and Biostatistics (F.D, B.A.C), Harvard T.H. Chan School of Public Health, Boston.

Abstract

Background: Associations between residential exposure to radon and harmful health outcomes were well-documented. However, causal modeling methods have never been used to evaluate the impacts of residential radon exposure.

Objectives: We estimated the causal effect of long-term residential radon exposure on mortality rate in Medicare beneficiaries in the New England region.

Methods: We calculated the all-cause mortality of Medicare beneficiaries for 1,240 Zip Code in New England from 2000 to 2015. We obtained 141,665 basement radon measurements from Spruce Environmental Technologies, Inc and estimated the annual ZIP Code-specific average basement radon level. The multi-panel nature of the data allowed us to estimate the impacts of residential radon with a difference-in-difference (DiD) experimental design. We applied a logistic regression model adjusted for age, gender, race, eligible for Medicaid, particulate matter, and temperature to estimate the percent increase in mortality associated with a unit increase in residential radon. To improve the plausibility of the DiD design assumption, we split the study population by county, restricted the analysis to each county, and then used a meta-analysis method to aggregate the county-specific effects. We performed a subgroup analysis to identify the population or region with a higher risk. Furthermore, we investigated the effects of residential radon on the mortality rate of beneficiaries with chronic underlying health issues.

Results: For each unit increase in annual average radon (1 pCi/L), the all-cause mortality rate in Medicare participants increases by 4.0% [95% confidence interval (CI): 0.9%, 7.2%]. The percentage increment of mortality in female participant is 5.0% (95% CI: 0.8%, 9.3%), higher than that in male participants (percent increase: 2.8% , 95% CI: -1.2%, 6.7%). Furthermore,

residential radon exposure is statistically associated with a significantly elevated mortality rate in the beneficiaries with underlying chronic cardiovascular disease (percent increase: 3.6%, 95% CI: 0.2%, 7.0%).

Conclusions: Residential exposure to radon is casually associated with a statistically significantly elevated all-cause mortality in the Medicare population in New England.

Introduction

Radon-222 is an odorless and chemically inert gaseous radioactive material with a half-life of 3.8 days. It arises naturally from the decay of Uranium-238, which is ubiquitous in the crust.(Otton 1992) Radon-222 (referred to as radon in this paper) then decays through a chain of short-lived and long-lived intermediates, to the stable Lead-206. Before decaying, radon gas can diffuse through the soil and get emanated into the atmosphere. It can also infiltrate into residential buildings through the openings in the foundation and water supply system. Radon concentration can build up to a high level in the building with a high radon infiltration rate and poor ventilation condition.(National Council on Radiation Protection and Measurements. 1984; "Radon in Homes: Report of the Council on Scientific Affairs, American Medical Association" 1987)

Residential exposure to radon is the leading cause of lung cancer for non-smokers in the U.S.("Health Effects of Radon Exposure: Report of the Council on Scientific Affairs, American Medical Association" 1991) This health hazard does not come primarily from radon itself, but rather from its radioactive progenies.(National Research Council 2006) The short-lived immediate progenies of radon quickly bond with the passing by water vapor and atmospheric gases and form ultrafine clusters. Part of these clusters then attach to airborne particles and become inhalable.(Porstendörfer 1994; Gründel and Porstendörfer 2004) After being inhaled, these radioactive progenies tend to be deposited on the bronchial epithelium, thus expose neighboring cells to high-energy α particles that induce the carcinogenesis process.(Darby et al. 2005; Duan et al. 2015)

Particulate-bound radon progenies are absorbed into the circulatory system, alimentary tract, and lymphatics.(Marsh and Bailey 2013) The long-lived progenies, including α-emitting Po-210,

could accumulate in the central nervous system and pulmonary tissues and thus could induce deleterious biological effects on tissues other than the respiratory tract.(Santos et al. 2020; Momcilovic et al. 2001) Blomberg et al. 2019 reported the modification effect of residential radon on the association between fine particulate matter ($PM_{2.5}$) and all-cause mortality in urban Medicare beneficiaries.(A. J. Blomberg et al. 2019) Yitshak-Sade et al. 2019 associated residential radon exposure with the mortality in Medicare beneficiaries residing in Mid-Atlantic and Northeastern U.S states.(Yitshak-Sade et al. 2019) However, Turner et al. 2012 found no clear associations between radon and nonrespiratory mortality in the Cancer Prevention Study II cohort.(Turner et al. 2012) Contradictory evidence reported by published studies highlight the importance of narrowing down the uncertainty in the association through a casual modeling method. Furthermore, existing studies regarding this topic are subject to limitations in the exposure assessment, including the coarse resolution of exposure assessment, a lack of temporal variation in the exposure.

We applied the Difference-in-Difference (DiD) experimental design, a quasi-experiment approach, to estimate the impacts of residential radon on all-cause mortality in all Medicare beneficiaries. We obtained residential radon measurements undertaken by a local radon test service provider. This data source allowed us to investigate the spatiotemporal variation of residential radon and predict residential exposure to radon at a higher resolution than previous studies.

Materials and Methods

Mortality Data

We obtained the Medicare beneficiary denominator file from the Center for Medicare and

Medicaid Services.(ResDac 2018) Our study population is composed of all Medicare beneficiaries 65 or older at enrollment and residing in the study area (Figure 3-1) for at least one year from 2000-2015. For each yearly record of a beneficiary, we extracted details including age, race, sex, Medicaid eligibility, residential ZIP Code, and date of death. The outcome of this study is the annual ZIP Code-level all-cause mortality rate in the Medicare beneficiaries. We also calculated the outcome in specific populations, including male/female subgroups, white/non-white subgroups, age-dependent subgroups, and Medicaid eligible/ineligible subgroups.

We obtained the emergency admission records of the beneficiaries enrolled in the Fee-For-Service program of Medicare. A participant was defined as having the specific chronic health issue after having an emergency admission for the causes (either primary cause or secondary cause), including chronic obstructive pulmonary diseases (COPD; ICD-9 490-496, except 493), cardiovascular diseases (CVD; ICD-9 390–459 and 410–414), congestive heart failure (CHF; ICD-9 428) and diabetes mellitus (DM; ICD-9 250). We calculated the annual mortality rates in the participants with these underlying health.

Residential Radon Data

We obtained radon measurements from Spruce Environmental Technologies, Inc, which detected the residential radon level in 8,9076 buildings in the New England region during our study period. We selected radon measurements detected by activated charcoal adsorption devices (AC) in the lowest livable level of each place under closed-house conditions. We did not have access to the street address information of clients. Instead, we used the 5-digit ZIP Code of the building as a proxy to the residential address. We subsequently estimated the monthly ZIP Code-level residential radon level based on the measurements conducted in the ZIP Code and month.

To approximate a random sampling of all buildings within the community, we only selected the measurements conducted before mitigation or during realty property transactions. We selected 7,6767 radon measurements from 5,8214 buildings (Figure 3-S1). We calculated 1,173 monthly ZIP Code average residential radon levels for 317 ZIP Codes with over five measurements in at least one month during our study period. We trained a Random Forest model(Breiman 2001) to predict the ZIP Code-specific monthly radon based on predictors including radon potential, the concentration of Uranium-238 in topsoil,(Joseph S. Duval, John M. Carson, Peter B. Holman 2005) the slope of the terrain, the granularity of ground surface material,(David R. Soller, Marith C. Reheis, Christopher P. Garrity 2009) soil moisture, air temperature, the height of planetary boundary layer, season and calendar year. We optimized the model through a grid-search method and evaluated the uncertainty of the model via ten-fold cross-validation. Finally, we predicted the monthly residential radon level for 1,240 ZIP Codes based on the fitted model (Figure 3-1).

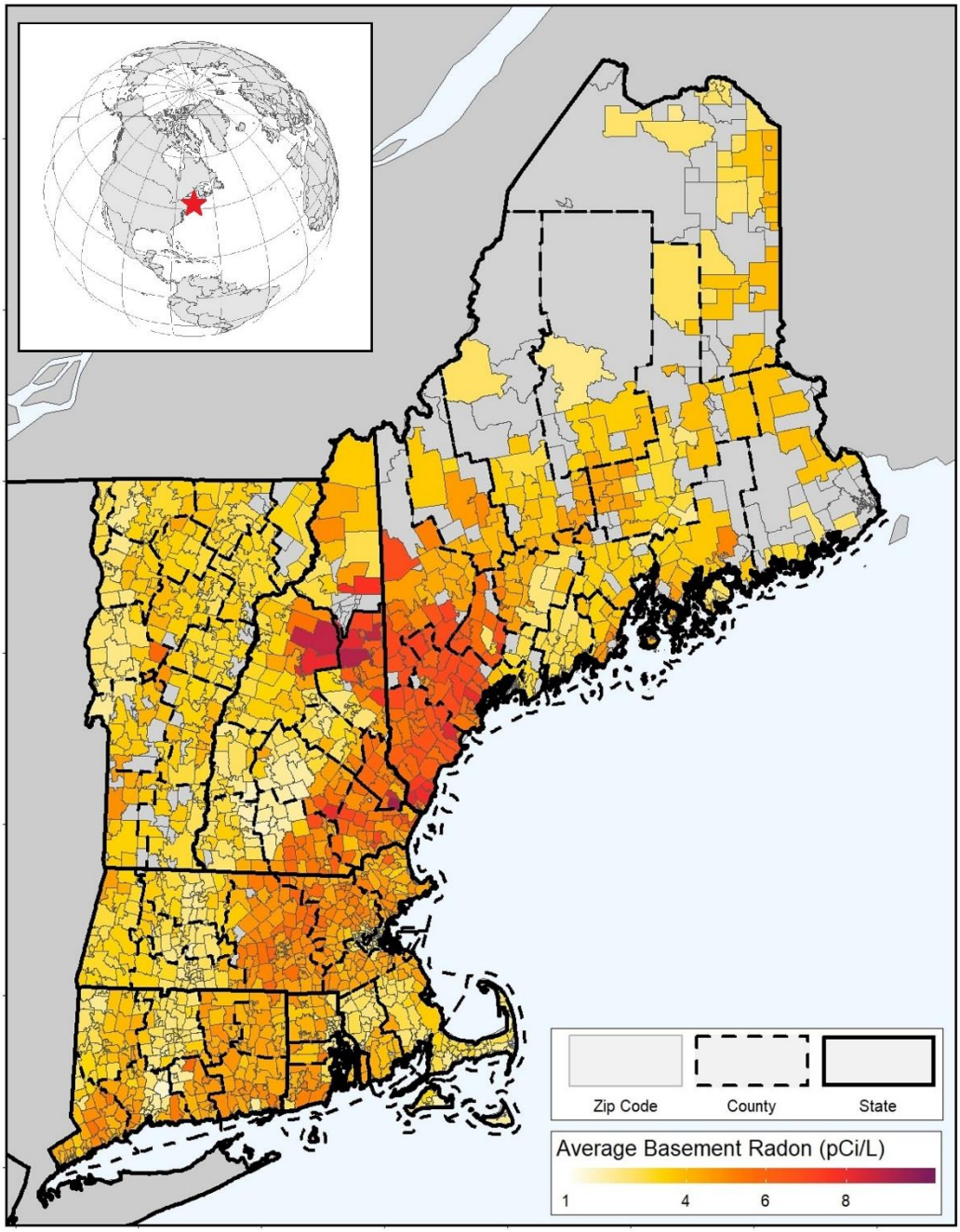


Figure 3-1 Study region and the average predicted Zip Code-specific* basement radon level from 2001 to 2015.

* Radon levels in the Zip Codes with fewer than three measurements were not predicted. Shown as light gray polygons, these ZIP Codes are mostly in northern Maine, where population density is low.

Potential Confounding Factors

We collected potential environmental confounding factors that associate mortality and residential radon simultaneously. Our 1 km×1 km grid-based fine particulate matter ($PM_{2.5}$) estimation was from a previously published model.(Di et al. 2016) The model predicts daily $PM_{2.5}$ concentration from remote sensing-based aerosol optical depth, land use information, and chemical transport model through a neural network algorithm. We first aggregated the daily prediction by year and then calculated the annual population-weighted average concentration of $PM_{2.5}$ for each ZIP Code.(Doxsey-Whitfield et al. 2015) We obtained the highest and lowest monthly temperatures for each ZIP Code and year from North America Regional Reanalysis dataset.(Mesinger et al. 2006)

We first obtained ZIP Code-level demographic and social-economic confounding factors by aggregating the individual-level information provided in mortality data. We then obtained ZIP Code-level potential social-economic confounding factors, including annual median household income, owner-occupied housing units median value, population percent below the poverty line, population percent without high school diplomas, and homeownership rate. These were obtained from the 2000 and 2010 US Census and the American Community Survey (ACS) and linearly extrapolated to account for the covariates' time-varying nature.

Analysis Methods

We used Difference-in-Difference (DiD) method,(Dimick and Ryan 2014; Wing, Simon, and Bello-Gomez 2018) a quasi-experimental design, to investigate the causal association between residential radon exposure and the mortality rate. A pair of health economic studies applied the

DiD approach to evaluate the effects of the 2011 Accreditation Council for Graduate Medical Education duty hour reforms.(Rajaram et al. 2014; Patel et al. 2014) The classic DiD method designed for dummy policy intervention can be generalized to estimate the effects of continuous treatment. Specially, we compared a treatment ZIP Code whose change in mortality and radon level during time ***t*** are $\Delta Mort_T$ and ΔRn_T with the contemporary control ZIP Code, whose variation in mortality and residential radon are $\Delta Mort_C$ and ΔRn_C, respectively. We can estimate the association between mortality rate and residential radon based on this two-by-two comparison as follows: $(\Delta Mort_T - \Delta Mort_C)/(\Delta Rn_T - \Delta Rn_C)$. The association is considered causal under the assumption that the trends of outcome in the treatment group and control group are the same in case they are exposed to the same level (referred to as parallel trend assumption).

Instead of permutating all two-by-two comparisons, we treated residential radon exposure as a fixed effect in a regression model to estimate the impacts based on the repeated cross-sectional dataset.(Wang et al. 2016) The fixed effect estimator enables us to estimate the uncertainty in the predicted effect. Additionally, we can use the generalized linear model to allow for a dependent variable whose residuals do not follow the normal distribution. Another benefit of employing the regression model includes the flexibility of adjusting for the observed potential confounding factor, which varies both spatially and temporarily. In the context of regression analysis, the parallel trend assumption is equivalent to the assumption that there is no unobserved confounder varying both across the region and over time. We used a logistic regression model with overdispersion to implement the DiD design as follows:

$$Logit(Y_{p,t}) = \beta_0 + \beta_1 Rn_{p,t} + \sum_{i=2\ldots5} \beta_i X_{i,p,t} + \sum_{j=6,7,8} \beta_j Z_{i,p,t} + \sum_{p \neq p_R} \beta_p I_p + \sum_{t \neq t_R} \beta_t I_t$$

where $Y_{p,t}$ is the annual ZIP Code-level all-cause mortality rate (count of death out of all persons

at the beginning of the year) in ZIP Code p at year t, $Rn_{p,t}$ represents the previously-estimated residential exposure to radon. $X_{2\ldots5,p,t}$ represents four ZIP Code-level demographic and social-economic factors obtained by aggregating individual-level data, including average age, percent of female beneficiaries, percent of white beneficiaries, and the percent of beneficiaries eligible for Medicaid. $Z_{6,\ldots8,p,t}$ represents three potential environmental confounding factors, including the concentration of $PM_{2.5}$, the highest monthly temperature in the summer, and the lowest monthly temperature in the winter; I_p and I_t represent indicators functions for each ZIP Code and year. In summary, the DiD approach controlled for four categories of factors, including 1) residential radon, the exposure of interest; 2) spatially-varying but temporally-invariant factors, adjusted for by the ZIP Code indicator function; 3) temporally-varying but spatially-invariant factors, adjusted for by the year indicator function; 4) observed potential confounding factors which vary spatiotemporally, represented by X and Z.

To make the estimated association a causal relationship, we have to assume the non-existence of spatiotemporally varying confounder because its variation can not fully be captured by I_p and I_t. Specifically, the assumption is violated under the condition that an unobserved confounder factor changes distinctly over time across ZIP Codes. This assumption is hard to hold because of the heterogeneity of demographic and environmental conditions of our study extent. For instance, winter temperature is a well-understood confounder of this association because of its negative association with mortality and positive correlation with basement radon.(“Radon in Homes: Report of the Council on Scientific Affairs, American Medical Association” 1987) The trend of winter temperature in the coastal area is not parallel with the corresponding trend in inland regions. However, this assumption is more plausible when the treatment and control ZIP Codes are from the same county because they share the same healthcare facilities, grocery chains, fast

food restaurants, tobacco retailers, and have a similar climate. Methodologically, we first split the study region by counties, then fitted the logistic regression model in each county, and finally used a random-effects meta-analysis method to aggregated the results from the county-level models.

We first associated mortality rate to residential radon in the whole study population. We then restricted the analysis to gender- and Medicaid eligibility-based subgroups to evaluate the health effects in diverse subgroups. We did not perform the analysis for ethnic minority subgroups due to a lack of power. Furthermore, we investigated the impact of radon in the subgroups with a history of chronic health conditions, including COPD, CVD, CHF, and DM. To evaluate the likelihood of violating the parallel trend assumption, we tested the sensitivity of our results to outside potential socio-economic confounders other than those originally included. The assumption does not hold if the estimated effects differ significantly between the original model and the modified model adjusting for outside social-economic potential confounders.

The analysis was conducted on the Cannon cluster, supported by the Research Computing Group, and on the Research Computing Environment, supported by the Institute for Quantitative Social Science, both at Harvard University, Faculty of Arts and Sciences. We used R software (version 3.4.2), stats package (version 3.6.2),(R Core Team 2017), and metaphor package (version 2.1.0)(Viechtbauer 2010) to perform the analysis.

Results

We studied 1,240 ZIP Codes in New England during 2000-2015. Our primary study population is 56.5% female, 96.5% white, and 11.0% dual-eligible for Medicaid. Out of the study participants, 10.7% had a history of COPD, 38.2% had a history of CVD, 11.6% had a history of

DM, and 11.4% had a history of CHF.

Table 3-1. Distribution of the Zip Code-specific exposure, outcomes, and the potential confounding factors among the 1752 Zip Code in New England from 2000 to 2015.

Variable	Mean	5th Percentile	25th Percentile	Median	75th Percentile	95th Percentile
Exposure: Basement Radon (pCi/L)	3.9	2.3	3.0	3.8	4.7	6.3
Outcome: Mortality rate (%)						
All participants	4.6%	2.9%	3.9%	4.6%	5.3%	6.5%
Male	4.8%	2.4%	3.8%	4.8%	5.7%	7.3%
Female	4.5%	2.4%	3.7%	4.5%	5.3%	6.7%
White	4.0%	2.9%	4.0%	4.7%	5.4%	6.5%
Non-white	7.5%	0.0%	2.9%	4.5%	7.7%	14.3%
Eligible for Medicaid	8.8%	0.0%	5.7%	8.1%	11.1%	17.9%
Ineligible for Medicaid	4.2%	2.5%	3.6%	4.2%	4.9%	6.0%
Age in [65, 75)	1.9%	0.6%	1.3%	1.8%	2.4%	3.4%
Age in [75, 85)	5.2%	2.5%	4.1%	5.1%	6.1%	8.1%
Age in [85, 95)	13.3%	6.7%	10.8%	13.0%	15.4%	20.8%
Chronic obstructive pulmonary disease (COPD)	20.5%	6.7%	15.4%	20.1%	25.0%	35.4%
Cardiovascular disease (CVD)	13.5%	7.1%	10.8%	13.0%	15.7%	21.7%
Congestive heart failure (CHF)	24.1%	9.5%	19.0%	23.8%	28.8%	39.4%
Diabetes mellitus (DM)	15.4%	4.0%	10.9%	14.9%	19.2%	28.6%
Potential Confounders						
PM2.5 ($\mu g/m^3$)	7.4	4.6	5.9	7.1	8.7	10.9
Summer Temperature (°C)	21.5	19.1	20.6	21.8	22.6	23.3
Winter Temperature (°C)	-4.4	-9.0	-6.4	-4.1	-2.5	0.2
White	97.0%	91.5%	96.8%	98.1%	99.0%	99.6%
Medicaid	10.6%	2.0%	4.8%	8.1%	13.6%	28.2%

Table 3-1 summarise the mortality rate in the whole study population and the sub-populations. The average mortality rate in our study population is 4.6% [Interquartile range (IQR): 3.9, 5.3%]. Female beneficiaries, white beneficiaries, and the participants non-eligible for Medicaid have lower mortality rates than their complementary subgroups, respectively. The average mortality

rates are higher in the sub-populations with a history of chronic disease, compared to the primary study population.

We evaluated the performance of the radon prediction model with a ten-fold cross-validation method (Figure 3-2). The standard deviation of the prediction residual is 1.43 pCi/L; the correlation (R^2) between the predicted and the measured is 0.68. There is no clear pattern of residual concerning the predicted radon level (Figure 3-2B), indicating that the potential exposure misclassification is independent of the actual exposure level, and thus is unable to bias our effect estimation, but would increase the confidence interval.

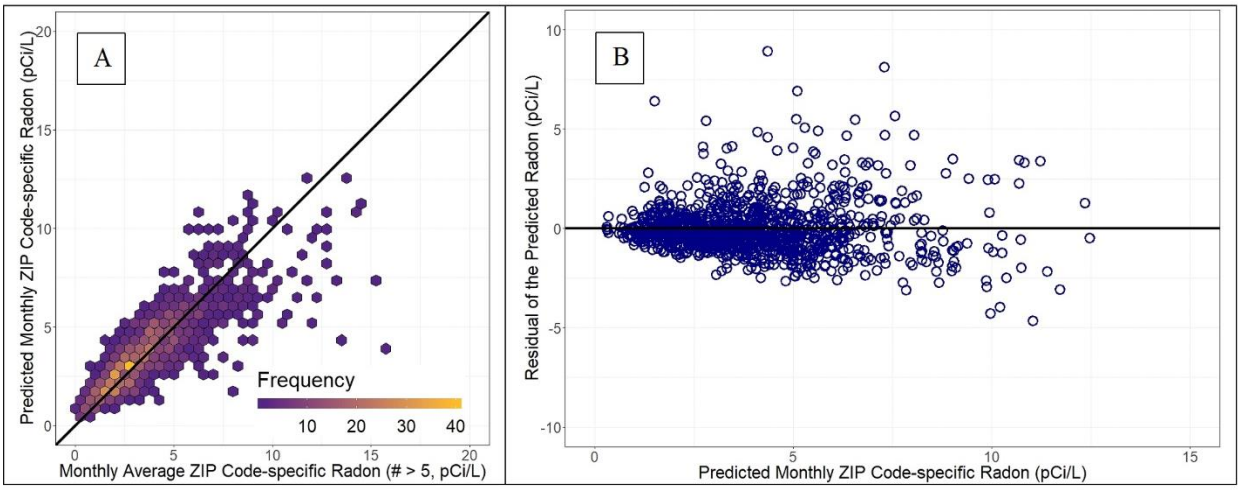


Figure 3-2. Diagnosis plots of the residential basement radon prediction model.

Based on this prediction model, the average radon level over the study period is 3.94 pCi/L, slightly lower than EPA's guideline level (4 pCi/L) for mitigation. As shown in Figure 3-1, ZIP Codes with high radon levels are in northern New Hampshire and southern Maine. The average residential radon decreased gradually during the study period (Figure 3-S1), probably due to the increasing prevalence of radon mitigation. From 2000 to 2015, the percent of ZIP Code, with >4 pCi/L radon level decreased from 46.6% to 39.2%.

Based on the meta-analysis of pooling county-level effects, a unit increase in residential

basement radon level was casually associated with a 4.0% increase (95% CI, 0.9,7.2%) in all-cause mortality in Medicare beneficiaries in New England during 2000-2015. By restricting the study population to female Medicare beneficiaries, we found a 5.0% increase (95% CI, 0.8, 9.3%) in mortality per unit increase in residential radon. Meanwhile, the impact on male beneficiaries is not statistically significant (the percent increase is 2.8%, 95% CI: -1.2, 6.8%). For white beneficiaries and the beneficiaries non-eligible for Medicaid, the percent changes in mortality are 4.9% (95% CI: 1.8, 8.0%) and 4.0 (95% CI: 0.0, 8.0%), respectively. We also investigated the state-level health effects by pooling the county-level impacts estimated for each county within the state (Figure.3). The state-specific associations are not statistically significant other than in Connecticut (percent change 4.3%, 95% CI: 0.4, 8.1%).

The percent change in mortality of a subgroup with a history of CVD is 3.6% (95% CI: 0.2, 7.0%). The associations between mortality and residential radon exposure are not statistically significant in subgroups with other chronic diseases. For participants with a history of COPD, DM, and CHF, the associated percent changes are 3.9% (95% CI: -2.7, 10.4%), 5.1% (95%CI: -1.1, 11.3%), and 0.6% (95% CI: -3.6%, 4.8%), respectively. A lack of power, jointly with the uncertainties of exposure assessment, probably caused the wide confidence intervals.

We evaluated the plausibility of the parallel trend assumption by testing the sensitivity of estimated effects to the potential social-economic confounders from Census datasets. For instance, ZIP Code-level median house value significantly associates with a lower mortality (R^2=0.24, p-value<0.01), and a higher residual radon level (R^2= 0.035, p-value<0.001). As summarized in Table.3, the estimated effects based on modified models are similar to that based on the original model. After adding median house value, the percent increase is 3.5% (95% CI: 0.9, 6.0%). We further evaluated the sensitivity of our estimated effects to the exclusion of

observed covariate. As summarized in Table 3-3, the estimated effects based on the deducted models are similar to the original model.

Table 3-2. The results of subgroup analysis

Subgroup	Percent Increase Estimation	95% Confidence Interval
All participants	4.0%	(0.9% ,7.2%)
Male	2.8%	(-1.2% ,6.8%)
Female	5.0%	(0.8% ,9.3%)
Non-white	0.5%	(-6.7% ,7.7%)
White	4.0%	(0.0% ,8.0%)
Non-eligible for Medicaid	4.9%	(1.8% ,8.0%)
Age in [65, 75)	6.3%	(-0.2% ,12.8%)
Age in [75, 85)	3.5%	(-1.6% ,8.6%)
Age in [85, 95)	-0.7%	(-5.5% ,4.1%)
Chronic obstructive pulmonary disease (COPD)	3.9%	(-2.7% ,10.4%)
Cardiovascular disease (CVD)	3.6%	(0.2% ,7.0%)
Diabetes mellitus (DM)	5.1%	(-1.1% ,11.3%)
Congestive heart failure (CHF)	0.6%	(-3.6% ,4.8%)
Connecticut	4.3%	(0.4% ,8.1%)
Maine	6.8%	(-6.6% ,20.2%)
Massachusetts	3.9%	(-0.5% ,8.3%)
New Hampshire	1.4%	(-4.0% ,6.9%)
Rhode Island	-1.6%	(-27.7% ,24.6%)
Vermont	8.6%	(-16.0% ,33.1%)

Discussion

In this study, we investigated the impact of long-term residential exposure to radon on all-cause mortality in Medicare beneficiaries in New England. After adjusting for the observed

confounders in the logistic regression model, we found a significant association between radon exposure and mortality rate on ZIP Code-level. To our best knowledge, this is the first study to estimate the causal health impacts of this ubiquitous environmental exposure.

We found a remarkable gender-dependent difference in the percent increase of mortality rate associated with a unit increase in radon. However, the two proportional hazards cannot be compared directly because of the different baseline mortality rate. After adjusting for the difference, the impact on female participants is still higher than that on male participants. This distinction can be explained by the longer hours female beneficiaries stay indoor, compared to male participants. Similarly, we found a higher proportional hazard in the subgroup with age in [65, 75) compared to the elder subgroup with age in [75,85). After adjusting for the difference in the baseline, the elder subgroup has a higher absolute hazard, relative to the younger subgroup.

In this study, we found that residential radon only significantly impacts the mortality in beneficiaries with a history of CVD (Table 3-2). The impacts on the participants with a history of COPD or a history of DM are positive, but not statistically significant. After adjusting for the baseline mortality in CVD patients (Table 3-1), the absolute impact associated with a unit increase in residential radon is approximately elevating the mortality rate from 13.5% to 14%, significantly larger than the impact on the whole study population. This result indicated that beneficiaries with CVD are more susceptible to radon effect.

Table 3-3. The increases in mortality casually associated with a unit increment of residential radon.

Model Modification	Percent Increase Estimation	95% Confidence Interval
Excluding existing terms		
Excluding Race	4.1%	(1.1% ,7.2%)
Excluding Medicaid	4.1%	(0.5% ,7.6%)
Excluding Age	4.0%	(0.8% ,7.3%)
Excluding Gender	4.5%	(1.5% ,7.6%)
Excluding $PM_{2.5}$	4.0%	(1.2% ,6.8%)
Excluding Summer Temperature	4.4%	(1.7% ,7.2%)
Excluding Winter Temperature	3.9%	(1.1% ,6.7%)
Adding SES confounders		
Adding Poverty	4.0%	(1.1% ,6.8%)
Adding Median House Value	3.5%	(0.9% ,6.0%)
Adding Mean Household Income	4.4%	(1.5% ,7.3%)
Adding Percent Cccupied by the Owner	3.9%	(1.0% ,6.8%)
Adding Percent Without High School Diploma	4.1%	(1.2% ,6.9%)

For comparison with the previously-published study, we first adjusted our percent increase in mortality to reflect a 0.7 pCi/L increase in exposure. We found a 2.8% (95% CI: 0.6, 5.0%) increase in all-cause mortality in Medicare beneficiaries of New England in consistence with that of 2.62% (95% CI: 2.52, 2.73%) found in the Medicare population living in the Mid-Atlantic and Northeastern U.S states by Yitshak-Sade et al. 2020. We then converted our percent increase to

reflect a 2.7 pCi/L (100 Bq/m^3) increase in radon concentration. We found 10.9% (95% CI: -7.1, 30.6%) increase in mortality of Medicare beneficiaries with COPD. By comparison, Turner et al. 2012 reported an increase of 13% (95% CI 5.0, 21%) in 1.2 million participants of the American Cancer Study.

We used the DiD method to simulate the counterfactual mortality, thus to investigate the causal link through the observational study. The validity of this simulation largely depends on the plausibility of the parallel trend assumption. Under this assumption, adding an extra spatiotemporally varying covariate in the model could not change the estimated effect remarkably. We first investigated the sensitivity to omitting an "observed confounder" that was calculated from the individual-level information provided in the mortality dataset. We then introduced outside social-economic confounders to serve as "unobserved" confounders and evaluated the sensitivity of the estimated effects to these potential confounders. The results of the sensitivity analysis (Table 3-3) suggest that our model is robust to be changed by unobserved confounding factors.

One major strength of this study is the exposure assessment. Previous studies concerning the health effects of residential radon commonly used the county mean radon levels that were modeled by Lawrence Berkeley National Laboratory partially based on the EPA/ State Residential Radon Survey short-term measurements conducted before 1992.(Price and Nero 1996) Although geological condition primarily determines the indoor radon level, meteorological factors, such as soil moisture, barometric pressure, and temperature, played essential roles by affecting the radon pressure difference between the basement and the underground.("Radon in Homes: Report of the Council on Scientific Affairs, American Medical Association" 1987) There is a nearly monotonic declining trend in residential radon levels across our study region,

partially due to an increasing prevalence of radon mitigation service and a wide-adoption of radon-resistance materials in the new construction. Ignoring the temporal trend in residential radon exposure induces non-random exposure misclassification, thus may bias the estimated association. Additional uncertainty is induced by assigning the same residential radon level to all residents living within the same county, regardless of the heterogeneity of geological conditions whose variation takes place at a smaller spatial scale.

We used the predicted ZIP Code-level annually average basement radon as a proxy to residential exposure to radon. Even though we collected diverse geological and meteorological covariates to train the random forest model, the R^2 of the model is 0.68, meaning over 30% of the variation was left unexplained. The pronounced within-ZIP Code variation partially causes low explaining power and induce additional uncertainties in the exposure measurement error. Furthermore, radon concentrations within a house decrease significantly from the basement to the living rooms and vary remarkably from room to room. Actual individual-level radon exposure also depends on the amount of time spent indoors. All these sources of uncertainties are theoretically independent of the actual level, and thus considered random exposure error. These random residuals may not bias the result but may enlarge the uncertainty in the estimated effects.

One limitation of this study is that we investigated the association of interest on the ZIP Code-level. To comply with the spatial resolution of exposure assessment, we aggregated individual-level death records to the ZIP Code-level mortality rate in an ecological form. However, an ecological model of aggregated events is technically equivalent to an individual-level model when the exposure is assigned to each individual in the unit.(Lu and Zeger 2007; Wang et al. 2016) Another limitation of this study is the lack of potential confounders concerning lifestyle behavior, such as smoking history and body weight index. It is reasonable to assume that our

estimated results are not sensitive to these factors because there are well-understood associations between lifestyle behavior risks and the social-economic factors adjusted for in our model.(Hiscock et al. 2012; Morgenstern, Sargent, and Hanewinkel 2009) Also, these factors can be considered time-variant only in a small region because the participants share the same facility and retailer, thus did not influence our estimated associations.

Conclusions

We found a causal association between residential exposure to radon and all-cause mortality rate in Medicare beneficiaries in New England. Also, beneficiaries with underlying chronic cardiovascular disease are more susceptible to the impacts.

Appendix: Supplementary Materials

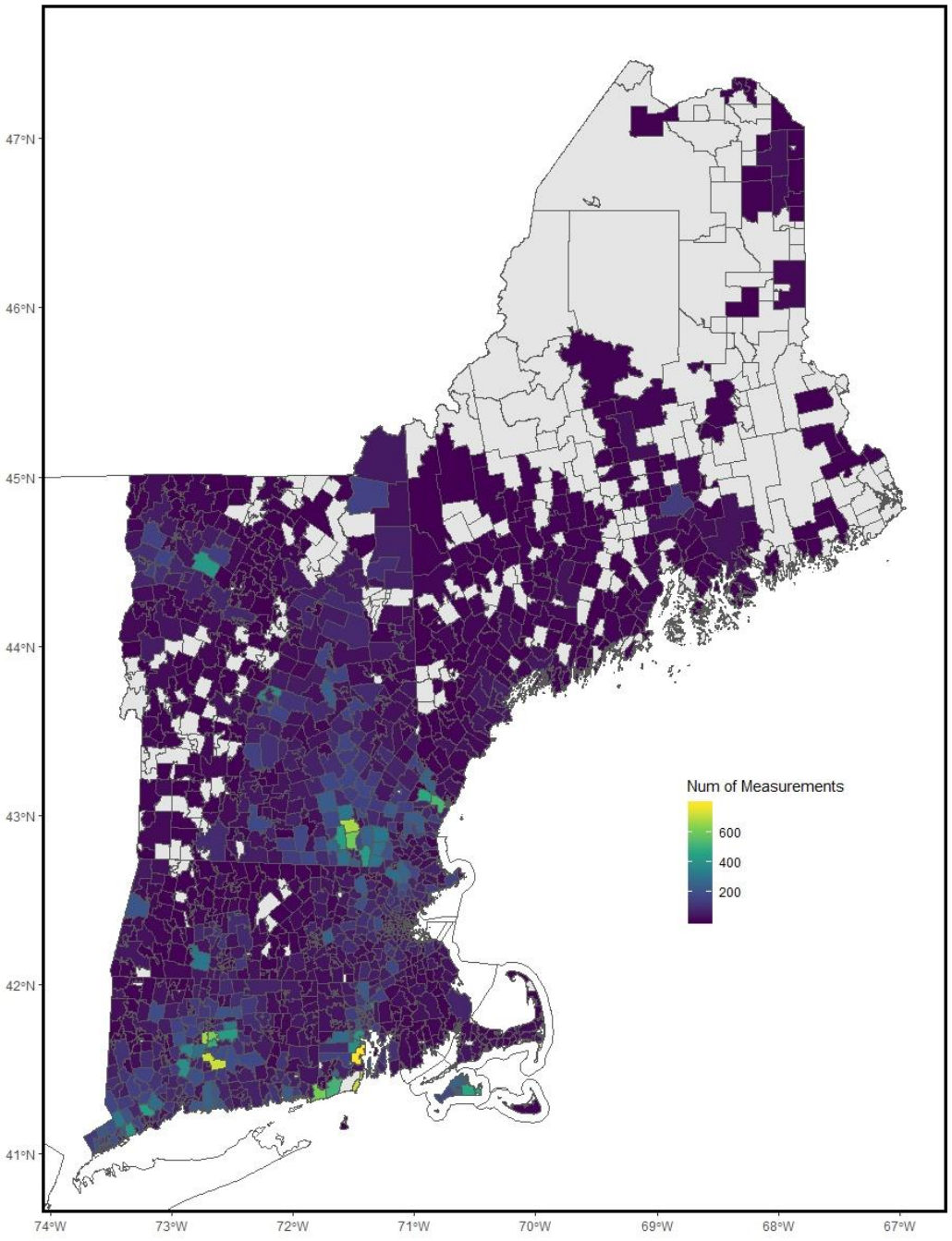


Figure 3-S1. The number of residential radon measurements undertaken by Spurce Environmetal Technologies, Inc in each ZIP Code in New England from 1996 to 2018.

Table 3-S1. The spatial-temporal variation of residential radon exposure (annually average and interquartile range of the ZIP Code-level radon, pCi/L).

Year	CT	ME	MA	NH	RI	VT
2000	3.90 (2.80,4.68)	4.63 (3.26,6.05)	3.93 (2.75,4.80)	4.60 (3.20,5.63)	3.80 (2.65,4.64)	3.35 (2.93,3.75)
2001	3.87 (2.80,4.63)	4.64 (3.21,6.05)	3.90 (2.73,4.76)	4.57 (3.14,5.68)	3.78 (2.63,4.62)	3.35 (2.93,3.70)
2002	3.89 (2.79,4.66)	4.72 (3.32,6.05)	4.01 (2.80,4.93)	4.63 (3.22,5.83)	3.84 (2.66,4.68)	3.43 (3.03,3.80)
2003	3.84 (2.74,4.60)	4.63 (3.20,5.94)	3.88 (2.70,4.70)	4.57 (3.22,5.58)	3.73 (2.56,4.58)	3.46 (3.08,3.85)
2004	3.91 (2.78,4.70)	4.63 (3.32,5.91)	3.94 (2.76,4.83)	4.55 (3.06,5.63)	3.77 (2.63,4.57)	3.41 (3.00,3.77)
2005	3.86 (2.80,4.61)	4.61 (3.33,5.91)	3.96 (2.82,4.83)	4.52 (3.04,5.62)	3.79 (2.66,4.53)	3.32 (2.93,3.66)
2006	3.84 (2.81,4.60)	4.57 (3.24,5.87)	3.91 (2.78,4.77)	4.47 (2.97,5.59)	3.79 (2.67,4.53)	3.30 (2.89,3.60)
2007	3.83 (2.77,4.56)	4.60 (3.35,5.81)	3.96 (2.80,4.91)	4.49 (2.98,5.66)	3.80 (2.64,4.53)	3.26 (2.95,3.54)
2008	3.85 (2.81,4.58)	4.57 (3.36,5.79)	3.91 (2.80,4.77)	4.49 (3.04,5.63)	3.78 (2.68,4.49)	3.27 (2.92,3.53)
2009	3.87 (2.86,4.57)	4.58 (3.35,5.79)	3.92 (2.84,4.77)	4.52 (3.08,5.65)	3.80 (2.76,4.50)	3.30 (3.02,3.56)
2010	3.85 (2.87,4.55)	4.63 (3.39,5.85)	3.97 (2.89,4.89)	4.48 (3.01,5.64)	3.80 (2.73,4.52)	3.27 (3.02,3.49)
2011	3.73 (2.79,4.41)	4.48 (3.35,5.54)	3.81 (2.81,4.64)	4.35 (3.03,5.40)	3.72 (2.77,4.50)	3.26 (2.97,3.48)
2012	3.66 (2.81,4.26)	4.44 (3.31,5.54)	3.78 (2.79,4.64)	4.30 (3.04,5.27)	3.65 (2.71,4.35)	3.26 (3.00,3.49)
2013	3.67 (2.86,4.29)	4.44 (3.32,5.54)	3.80 (2.82,4.65)	4.27 (3.04,5.24)	3.72 (2.80,4.43)	3.27 (3.01,3.53)
2014	3.63 (2.82,4.25)	4.39 (3.29,5.40)	3.79 (2.80,4.66)	4.25 (2.99,5.21)	3.67 (2.74,4.39)	3.20 (2.98,3.40)
2015	3.59 (2.79,4.19)	4.39 (3.35,5.45)	3.75 (2.76,4.60)	4.20 (2.95,5.18)	3.64 (2.73,4.36)	3.15 (2.94,3.36)

CONCLUSIONS

In this dissertation, we investigated the links among UOGD, ambient particle radioactivity, and health. We started by estimating the impact of UOGD on all-cause mortality. Subsequently, we investigated the contribution of extensive UOGD on the radioactivity of ambient particulate matter that could serve as an overlooked exposure pathway. Finally, we estimated the causal association of radon, the primary source of residential radioactivity exposure, to all-cause mortality.

In our first study, we found a significantly elevated risk of mortality for the medicare beneficiaries residing downwind of the UOGD site, compared to those living upwind, after adjusting the proximity. This association is considered causal because of the independence between wind direction and drilling activities. In our second study, we found a significant contribution of UOGD activities to local ambient particle radioactivity levels. In our third study, we found a significant association between residential radon exposure and all-cause mortality. The results of the second and third study jointly suggest an overlooked UOGD exposure pathway

BIBLIOGRAPHY

Allen, David T. 2014. "Atmospheric Emissions and Air Quality Impacts from Natural Gas Production and Use." *Annual Review of Chemical and Biomolecular Engineering* 5 (1): 55–75. https://doi.org/10.1146/annurev-chembioeng-060713-035938.

Andersen, P K, and R D Gill. 1982. "Cox's Regression Model for Counting Processes: A Large Sample Study." *The Annals of Statistics* 10 (4): 1100–1120.

Baskaran, M. 2011. "Po-210 and Pb-210 as Atmospheric Tracers and Global Atmospheric Pb-210 Fallout: A Review." *Journal of Environmental Radioactivity* 102 (5): 500–513. https://doi.org/10.1016/J.JENVRAD.2010.10.007.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using **Lme4**." *Journal of Statistical Software* 67 (1): 1–48. https://doi.org/10.18637/jss.v067.i01.

Blair, Benjamin D., Stephen Brindley, Eero Dinkeloo, Lisa M. McKenzie, and John L. Adgate. 2018. "Residential Noise from Nearby Oil and Gas Well Construction and Drilling." *Journal of Exposure Science and Environmental Epidemiology* 28 (6): 538–47.

Blomberg, Annelise J., Brent A. Coull, Iny Jhun, Carolina L.Z. Vieira, Antonella Zanobetti, Eric Garshick, Joel Schwartz, and Petros Koutrakis. 2019. "Effect Modification of Ambient Particle Mortality by Radon: A Time Series Analysis in 108 U.S. Cities." *Journal of the Air and Waste Management Association* 69 (3): 266–76. https://doi.org/10.1080/10962247.2018.1523071.

Blomberg, Annelise, Marguerite Nyhan, Marie-Abèle Bind, Pantel Vokonas, Brent Coull, Joel Schwartz, and Petros Koutrakis. 2020. "The Role of Ambient Particle Radioactivity in Inflammation and Endothelial Function in an Elderly Cohort." *Epidemiology* 31 (4).

Boudet, Hilary S., Chad M. Zanocco, Peter D. Howe, and Christopher E. Clarke. 2018. "The Effect of Geographic Proximity to Unconventional Oil and Gas Development on Public Support for Hydraulic Fracturing." *Risk Analysis* 38 (9): 1871–90. https://doi.org/10.1111/risa.12989.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45: 5–32.

Brown, Valeria J. 2014. "Radionuclides in Fracking Wastewater: Managing a Toxic Blend." *Environmental Health Perspectives* 122 (2): A50-5. https://doi.org/10.1289/ehp.122-A50.

Cabello, María, Concepción Dueñas, Esperanza Liger, Elisa Gordo, and Sergio Cañete. 2018. "Variables Influencing the Gross Alpha and Gross Beta Activities in Airborne Particulate Samples in Málaga, Spain." *Journal of Radioanalytical and Nuclear Chemistry* 315 (2): 299–307. https://doi.org/10.1007/s10967-017-5674-3.

Casey, Joan A., Elizabeth L. Ogburn, Sara G. Rasmussen, Jennifer K. Irving, Jonathan Pollak, Paul A. Locke, and Brian S. Schwartz. 2015. "Predictors of Indoor Radon Concentrations in Pennsylvania, 1989–2013." *Environmental Health Perspectives* 123 (11): 1130–37. https://doi.org/10.1289/ehp.1409014.

Casey, Joan A, David A Savitz, Sara G Rasmussen, Elizabeth L Ogburn, Jonathan Pollak, Dione G Mercer, and Brian S Schwartz. 2016. "Unconventional Natural Gas Development and Birth Outcomes in Pennsylvania, USA." *Epidemiology (Cambridge, Mass.)* 27 (2): 163–72. https://doi.org/10.1097/EDE.0000000000000387.

CDC (Center for Disease Control and Prevention). 2013. "Behavior Risk Factor Surveillance System. BRFSS 2013 Survey Data and Documentation." 2013. https://www.cdc.gov/brfss/annual_data/annual_2013.html.

Cheadle, L. C., S. J. Oltmans, G. Petron, R. C. Schnell, E. J. Mattson, S. C. Herndon, A. M. Thompson, D. R. Blake, and A. McClure-Begley. 2017. "Surface Ozone in the Colorado Northern Front Range and the Influence of Oil and Gas Development during FRAPPE/DISCOVER-AQ in Summer 2014." *Elem Sci Anth* 5 (0): 61. https://doi.org/10.1525/elementa.254.

Commission, U S Atomic Energy. 1961. "Geology and Geochemistry of Uranium in Marine Black Shales A Review Geology and Geochemistry of Uranium in Marine Black Shales A Review." Washington, D.C. https://pubs.usgs.gov/pp/0356c/report.pdf.

Cordeiro, Cristina, Paulo J.C. Favas, João Pratas, Santosh Kumar Sarkar, and Perumal Venkatachalam. 2016. "Uranium Accumulation in Aquatic Macrophytes in an Uraniferous Region: Relevance to Natural Attenuation." *Chemosphere* 156 (August): 76–87. https://doi.org/10.1016/J.CHEMOSPHERE.2016.04.105.

Czolowski, Eliza D., Renee L. Santoro, Tanja Srebotnjak, and Seth B.C. Shonkoff. 2017. "Toward Consistent Methodology to Quantify Populations in Proximity to Oil and Gas Development: A National Spatial Analysis and Review." *Environmental Health Perspectives* 125 (8): 086004. https://doi.org/10.1289/EHP1535.

Darby, S., D. Hill, A. Auvinen, J. M. Barros-Dios, H. Baysson, F. Bochicchio, H. Deo, et al. 2005. "Radon in Homes and Risk of Lung Cancer: Collaborative Analysis of Individual Data from 13 European Case-Control Studies." *British Medical Journal* 330 (7485): 223–26. https://doi.org/10.1136/bmj.38308.477650.63.

David R. Soller, Marith C. Reheis, Christopher P. Garrity, and D.R. Van Sistine. 2009. *Map Database for Surficial Materials in the Conterminous United States: U.S. Geological Survey Data Series 425, Scale 1:5,000,000*. Reston, VA: U.S. Geological Survey.

Di, Qian, Itai Kloog, Petros Koutrakis, Alexei Lyapustin, Yujie Wang, and Joel Schwartz. 2016. "Assessing PM 2.5 Exposures with High Spatiotemporal Resolution across the Continental United States." *Environmental Science & Technology* 50 (9): 4712–21. https://doi.org/10.1021/acs.est.5b06121.

Dimick, Justin B., and Andrew M. Ryan. 2014. "Methods for Evaluating Changes in Health Care Policy: The Difference-in-Differences Approach." *JAMA - Journal of the American Medical Association* 312 (22): 2401–2. https://doi.org/10.1001/jama.2014.16153.

Dominici, Francesca, Michael Greenstone, and Cass R. Sunstein. 2014. "Particulate Matter Matters." *Science*. American Association for the Advancement of Science. https://doi.org/10.1126/science.1247348.

Doxsey-Whitfield, Erin, Kytt MacManus, Susana B. Adamo, Linda Pistolesi, John Squires, Olena Borkovska, and Sandra R. Baptista. 2015. "Taking Advantage of the Improved Availability of Census Data: A First Look at the Gridded Population of the World, Version 4." *Papers in Applied Geography* 1 (3): 226–34. https://doi.org/10.1080/23754931.2015.1014272.

Duan, Peng, Chao Quan, Chunhui Hu, Jicai Zhang, Fei Xie, Xiuxue Hu, Zongtao Yu, et al. 2015. "Nonlinear Dose-Response Relationship between Radon Exposure and the Risk of Lung Cancer: Evidence from a Meta-Analysis of Published Observational Studies." *European Journal of Cancer Prevention*. Lippincott Williams and Wilkins. https://doi.org/10.1097/CEJ.0000000000000066.

Earth Resources Observation and Science (EROS) Center. 2012. "The National Land Cover Database." Reston, Virginia.

Eitrheim, Eric S., Dustin May, Tori Z. Forbes, and Andrew W. Nelson. 2016. "Disequilibrium of Naturally Occurring Radioactive Materials (NORM) in Drill Cuttings from a Horizontal Drilling Operation." *Environmental Science & Technology Letters* 3 (12): 425–29. https://doi.org/10.1021/acs.estlett.6b00439.

Enverus. 2019. "Enverus Drillinginfo Direct Access Application Programming Interface." 2019. https://app.drillinginfo.com/direct/.

Fisher, R Stephen. 1998. "Geologic and Geochemical Controls on Naturally Occurring Radioactive Materials (NORM) in Produced Water from Oil, Gas, and Geothermal Operations." *Environmental Geosciences* 5 (3): 139–50.

Fraass, Ronald. 2015. "RadNet National Air Monitoring Program." In *Nuclear Terrorism and National Preparedness*, 117–23. Springer, Dordrecht. https://doi.org/10.1007/978-94-017-9891-4_11.

Franklin, Meredith, Khang Chau, Lara J. Cushing, and Jill E. Johnston. 2019. "Characterizing Flaring from Unconventional Oil and Gas Operations in South Texas Using Satellite Observations." *Environmental Science & Technology* 53 (4): 2220–28. https://doi.org/10.1021/acs.est.8b05355.

Gründel, M., and J. Porstendörfer. 2004. "Differences between the Activity Size Distributions of the Different Natural Radionuclide Aerosols in Outdoor Air." *Atmospheric Environment* 38 (22): 3723–28. https://doi.org/10.1016/J.ATMOSENV.2004.01.043.

Health Effects Institute-Energy (HEI-Energy) Research Committee. 2019a. "Human Exposure To Unconventionl Oil and Gas Development: A Literature Survery For Research Planning (Draft For Public Comment)." Boston, MA.

———. 2019b. "Potential Human Health Effects Associated With Unconventional Oil and Gas Development : A Systematic Review Of The Epidemiology Literature." Boston, MA.

"Health Effects of Radon Exposure: Report of the Council on Scientific Affairs, American Medical Association." 1991. *Archives of Internal Medicine* 151 (4): 674–77. https://doi.org/10.1001/archinte.1991.00400040028007.

Hernández, F., J. Hernández-Armas, A. Catalán, J.C. Fernández-Aldecoa, and L. Karlsson. 2005. "Gross Alpha, Gross Beta Activities and Gamma Emitting Radionuclides Composition of Airborne Particulate Samples in an Oceanic Island." *Atmospheric Environment* 39 (22): 4057–66. https://doi.org/10.1016/J.ATMOSENV.2005.03.035.

Hill, Elaine, and Lala Ma. n.d. "Shale Gas Development and Drinking Water Quality." *American Economic Review: Papers & Proceedings* 2017 (5): 522–25. https://doi.org/10.1257/aer.p20171133.

Hiscock, Rosemary, Linda Bauld, Amanda Amos, Jennifer A. Fidler, and Marcus Munafò. 2012. "Socioeconomic Status and Smoking: A Review." *Annals of the New York Academy of Sciences*. Blackwell Publishing Inc. https://doi.org/10.1111/j.1749-6632.2011.06202.x.

Joseph S. Duval, John M. Carson, Peter B. Holman, and Arthur G. Darnley. 2005. "Terrestrial Radioactivity and Gamma-Ray Exposure in the United States and Canada: U.S. Geological Survey Open-File Report 2005-1413." https://pubs.usgs.gov/of/2005/1413/.

Kelsey, Timothy W., Mark D. Partridge, and Nancy E. White. 2016. "Unconventional Gas and Oil Development in the United States: Economic Experience and Policy Issues." *Applied Economic Perspectives and Policy* 38 (2): 191–214.

Koehler, Kirsten, J Hugh Ellis, Joan A Casey, David Manthos, Karen Bandeen-Roche, Rutherford Platt, and Brian S Schwartz. 2018. "Exposure Assessment Using Secondary Data Sources in Unconventional Natural Gas Development and Health Studies." *Cite This: Environ. Sci. Technol* 52: 6061–69. https://doi.org/10.1021/acs.est.8b00507.

Kolb, W.A., and M. Wojcik. 1985. "Enhanced Radioactivity Due to Natural Oil and Gas Production and Related Radiological Problems." *Science of The Total Environment* 45 (October): 77–84. https://doi.org/10.1016/0048-9697(85)90206-2.

Landrigan, Philip J., Howard Frumkin, and Brita E. Lundberg. 2020. "The False Promise of Natural Gas." *New England Journal of Medicine*. https://doi.org/10.1056/NEJMp1913663.

Lauer, Nancy E., Jennifer S. Harkness, and Avner Vengosh. 2016. "Brine Spills Associated with Unconventional Oil Development in North Dakota." *Environmental Science & Technology* 50 (10): 5389–97. https://doi.org/10.1021/acs.est.5b06349.

Lauer, Nancy E, Nathaniel R Warner, and Avner Vengosh. 2018. "Sources of Radium Accumulation in Stream Sediments near Disposal Sites in Pennsylvania: Implications for Disposal of Conventional Oil and Gas Wastewater." *Environmental Science & Technology* 52 (3): 955–62. https://doi.org/10.1021/acs.est.7b04952.

Lee, Eric W., L. J. Wei, David A. Amato, and Sue Leurgans. 1992. "Cox-Type Regression Analysis for Large Numbers of Small Groups of Correlated Failure Time Observations." In *Survival Analysis: State of the Art*, 237–47. Springer Netherlands. https://doi.org/10.1007/978-94-015-7983-4_14.

Li, Wenyuan, Marguerite M. Nyhan, Elissa H. Wilker, Carolina L.Z. Vieira, Honghuang Lin, Joel D. Schwartz, Diane R. Gold, et al. 2018. "Recent Exposure to Particle Radioactivity and Biomarkers of Oxidative Stress and Inflammation: The Framingham Heart Study."

Environment International 121 (December): 1210–16. https://doi.org/10.1016/J.ENVINT.2018.10.039.

Lu, Yun, and Scott L Zeger. 2007. "On the Equivalence of Case-Crossover and Time Series Methods in Environmental Epidemiology." *Biostatistics (Oxford, England)* 8 (2): 337–44. https://doi.org/10.1093/biostatistics/kxl013.

Marsh, J W, and M R Bailey. 2013. "A Review of Lung-to-Blood Absorption Rates for Radon Progeny." *Radiation Protection Dosimetry* 157 (4): 499–514. https://doi.org/10.1093/rpd/nct179.

McKenzie, Lisa M., James Crooks, Jennifer L. Peel, Benjamin D. Blair, Stephen Brindley, William B. Allshouse, Stephanie Malin, and John L. Adgate. 2019. "Relationships between Indicators of Cardiovascular Disease and Intensity of Oil and Natural Gas Activity in Northeastern Colorado." *Environmental Research* 170 (March): 56–64. https://doi.org/10.1016/j.envres.2018.12.004.

Mesinger, Fedor, Geoff DiMego, Eugenia Kalnay, Kenneth Mitchell, Perry C. Shafran, Wesley Ebisuzaki, Dušan Jović, et al. 2006. "North American Regional Reanalysis." *Bulletin of the American Meteorological Society* 87 (3): 343–60. https://doi.org/10.1175/BAMS-87-3-343.

Mohery, M., A. M. Abdallah, Z. M. Al-Amoudi, and S. S. Baz. 2014. "Activity Size Distribution of Some Natural Radionuclides." *Radiation Protection Dosimetry* 158 (4): 435–41. https://doi.org/10.1093/rpd/nct250.

Momcilovic, Berislav, H. A. Alkhatib, J. A. Duerre, M. Cooley, W. M. Long, T. R. Harris, and G. I. Lykken. 2001. "Environmental Lead-210 and Bismuth-210 Accrue Selectively in the Brain Proteins in Alzheimer Disease and Brain Lipids in Parkinson Disease." In *Alzheimer Disease and Associated Disorders*, 15:106–15. https://doi.org/10.1097/00002093-200104000-00012.

Morgenstern, Matthis, James D. Sargent, and Reiner Hanewinkel. 2009. "Relation between Socioeconomic Status and Body Mass Index: Evidence of an Indirect Path via Television Use." *Archives of Pediatrics and Adolescent Medicine* 163 (8): 731–38. https://doi.org/10.1001/archpediatrics.2009.78.

National Council on Radiation Protection and Measurements. 1984. *Evaluation of Occupational and Environmental Exposures to Radon and Radon Daughters in the United States*. Bethesda, MD: NCRP.

National Research Council. 2006. *Health Risks from Exposure to Low Levels of Ionizing Radiation: BEIR VII Phase 2*. Washington D.C: The National Academies Press.

Nyhan, Marguerite M., Brent A. Coull, Annelise J. Blomberg, Carol L.Z. Vieira, Eric Garshick, Abdulaziz Aba, Pantel Vokonas, Diane R. Gold, Joel Schwartz, and Petros Koutrakis. 2018. "Associations Between Ambient Particle Radioactivity and Blood Pressure: The NAS (Normative Aging Study)." *Journal of the American Heart Association* 7 (6): e008245. https://doi.org/10.1161/JAHA.117.008245.

Nyhan, Marguerite M., Mary Rice, Annelise Blomberg, Brent A. Coull, Eric Garshick, Pantel

Vokonas, Joel Schwartz, Diane R. Gold, and Petros Koutrakis. 2019. “Associations between Ambient Particle Radioactivity and Lung Function.” *Environment International* 130 (September): 104795. https://doi.org/10.1016/J.ENVINT.2019.04.066.

Olmstead, Sheila M., Lucija A. Muehlenbachs, Jhih Shyang Shih, Ziyan Chu, and Alan J. Krupnick. 2013. “Shale Gas Development Impacts on Surface Water Quality in Pennsylvania.” *Proceedings of the National Academy of Sciences of the United States of America* 110 (13): 4962–67. https://doi.org/10.1073/pnas.1213871110.

Otton, James K. 1992. *The Geology of Radon*. Washington, D.C: United States Geological Survey.

Patel, Mitesh S., Kevin G. Volpp, Dylan S. Small, Alexander S. Hill, Orit Even-Shoshan, Lisa Rosenbaum, Richard N. Ross, Lisa Bellini, Jingsan Zhu, and Jeffrey H. Silber. 2014. “Association of the 2011 ACGME Resident Duty Hour Reforms with Mortality and Readmissions among Hospitalized Medicare Patients.” *JAMA - Journal of the American Medical Association* 312 (22): 2364–73. https://doi.org/10.1001/jama.2014.15273.

Pennsylvania Department of Environmental Protection. 2016. “Technologically Enhanced Naturally Occuring Radioactive Materials (TENORM) Study Report.” Harrisburg,PA.

Perry, Simona L. 2013. “Using Ethnography to Monitor the Community Health Implications of Onshore Unconventional Oil and Gas Developments: Examples from Pennsylvania’s Marcellus Shale.” *NEW SOLUTIONS: A Journal of Environmental and Occupational Health Policy* 23 (1): 33–53. https://doi.org/10.2190/NS.23.1.d.

Porstendörfer, J. 1994. “Properties and Behaviour of Radon and Thoron and Their Decay Products in the Air.” *Journal of Aerosol Science* 25 (2): 219–63. https://doi.org/10.1016/0021-8502(94)90077-9.

Price, Phillip N, and Anthony V Nero. 1996. “Mapping of Mean Radon Concentrations, Using Survey Data and Covariates.” In *International Radon Symposium*. Berkley,CA.

R Core Team. 2017. “R: A Language and Environment for Statistical Computing.” Vienna, Austria.

“Radon in Homes: Report of the Council on Scientific Affairs, American Medical Association.” 1987. *JAMA: The Journal of the American Medical Association* 258 (5): 668–72. https://doi.org/10.1001/jama.1987.03400050110038.

Rajaram, Ravi, Jeanette W. Chung, Andrew T. Jones, Mark E. Cohen, Allison R. Dahlke, Clifford Y. Ko, John L. Tarpley, Frank R. Lewis, David B. Hoyt, and Karl Y. Bilimoria. 2014. “Association of the 2011 ACGME Resident Duty Hour Reform with General Surgery Patient Outcomes and with Resident Examination Performance.” *JAMA - Journal of the American Medical Association* 312 (22): 2374–84. https://doi.org/10.1001/jama.2014.15277.

Rasmussen, Sara G., Elizabeth L. Ogburn, Meredith McCormack, Joan A. Casey, Karen Bandeen-Roche, Dione G. Mercer, and Brian S. Schwartz. 2016. “Association Between Unconventional Natural Gas Development in the Marcellus Shale and Asthma

Exacerbations." *JAMA Internal Medicine* 176 (9): 1334. https://doi.org/10.1001/jamainternmed.2016.2436.

ResDac. 2018. "Master Beneficiary Summary File (MBSF) Base." Resdac.Org. 2018. https://www.resdac.org/cms-data/files/mbsf-base.

Rich, Alisa L, and Ernest C Crosby. 2013. "Analysis of Reserve Pit Sludge from Unconventional Natural Gas Hydraulic Fracturing and Drilling Operations for the Presence of Technologically Enhanced Naturally Occurring Radioactive Material (TENORM)." *NEW SOLUTIONS* 23 (1): 117–35. https://doi.org/10.2190/NS.23.1.h.

Royal Observatory of Belgium. 2019. "Sunspot Data from the World Data Center SILSO." https://doi.org/http://www.sidc.be/SILSO.

Santos, Nathalia Villa dos, Carolina Leticia Zilli Vieira, Paulo Hilario Nascimento Saldiva, Barbara Paci Mazzilli, Mitiko Saiki, Catia Heloisa Saueia, Carmen Diva Saldiva De André, Lisie Tocci Justo, Marcelo Bessa Nisti, and Petros Koutrakis. 2020. "Levels of Polonium-210 in Brain and Pulmonary Tissues: Preliminary Study in Autopsies Conducted in the City of Sao Paulo, Brazil." *Scientific Reports* 10 (1): 1–7. https://doi.org/10.1038/s41598-019-56973-z.

Schwartz, Joel D., Yan Wang, Itai Kloog, Ma'Ayan Yitshak-Sade, Francesca Dominici, and Antonella Zanobetti. 2018. "Estimating the Effects of PM2.5 on Life Expectancy Using Causal Modeling Methods." *Environmental Health Perspectives* 126 (12). https://doi.org/10.1289/EHP3130.

Stein, A. F., R. R. Draxler, G. D. Rolph, B. J. B. Stunder, M. D. Cohen, F. Ngan, A. F. Stein, et al. 2015. "NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System." *Bulletin of the American Meteorological Society* 96 (12): 2059–77. https://doi.org/10.1175/BAMS-D-14-00110.1.

Therneau, Terry M. 2019. "A Package for Survival Analysis in S." https://cran.r-project.org/package=survival.

Torres, Luisa, Om Prakash Yadav, and Eakalak Khan. 2018. "Risk Assessment of Human Exposure to Ra-226 in Oil Produced Water from the Bakken Shale." *Science of the Total Environment* 626: 867–74.

Turner, Michelle C, Daniel Krewski, Yue Chen, C Arden Pope, Susan M Gapstur, and Michael J Thun. 2012. "Radon and Nonrespiratory Mortality in the American Cancer Society Cohort." *American Journal of Epidemiology* 176 (9): 808–14. https://doi.org/10.1093/aje/kws198.

U.S. Energy Information Administration (EIA). 2019a. "Drilling Productivity Report." 2019. https://www.eia.gov/petroleum/drilling/.

———. 2019b. "Horizontally Drilled Wells Dominate U.S. Tight Formation Production." 2019. https://www.eia.gov/todayinenergy/detail.php?id=39752.

———. 2019c. "The Distribution of U.S. Oil and Natural Gas Wells by Production Rate." Washington, D.C.

U.S. Enverionmental Preotection Agency (EPA). 2016. “Hydraulic Fracturing For Oil And Gas: Impacts From The Hydraulic Fracturing Water Cycle On Drinking Water Resources In The United States (Final Report).” Washington D.C.

U.S. Environmental Protection Agency (EPA). 2012. “Expansion and Upgrade of the RadNet Air Monitoring Network: Conceptual Plan and Implementation Process. Office of Radiation and Indoor Air. Vol. I, 2012.”

Viechtbauer, Wolfgang. 2010. “Conducting Meta-Analyses in R with the Metafor Package.” *Journal of Statistical Software* 36 (3).

Wang, Yan, Itai Kloog, Brent A. Coull, Anna Kosheleva, Antonella Zanobetti, and Joel D. Schwartz. 2016. “Estimating Causal Effects of Long-Term PM2.5 Exposure on Mortality in New Jersey.” *Environmental Health Perspectives* 124 (8): 1182–88. https://doi.org/10.1289/ehp.1409671.

Wing, Coady, Kosali Simon, and Ricardo A. Bello-Gomez. 2018. “Designing Difference in Difference Studies: Best Practices for Public Health Policy Research.” *Annual Review of Public Health* 39 (1): 453–69. https://doi.org/10.1146/annurev-publhealth-040617-013507.

Xu, Yanqing, Mounika Sajja, and Ashok Kumar. 2019. “Impact of the Hydraulic Fracturing on Indoor Radon Concentrations in Ohio: A Multilevel Modeling Approach.” *Frontiers in Public Health* 7 (April): 76. https://doi.org/10.3389/fpubh.2019.00076.

Yitshak-Sade, Maayan, Annelise J. Blomberg, Antonella Zanobetti, Joel D. Schwartz, Brent A. Coull, Itai Kloog, Francesca Dominici, and Petros Koutrakis. 2019. “County-Level Radon Exposure and All-Cause Mortality Risk among Medicare Beneficiaries.” *Environment International* 130 (September): 104865. https://doi.org/10.1016/j.envint.2019.05.059.

Zhang, Tieyuan, Richard W. Hammack, and Radisav D. Vidic. 2015. “Fate of Radium in Marcellus Shale Flowback Water Impoundments and Assessment of Associated Health Risks.” *Environmental Science & Technology* 49 (15): 9347–54. https://doi.org/10.1021/acs.est.5b01393.