ATTACHMENT C

STUDY 12

Preventive Medicine Reports 8 (2017) 112–115

Contents lists available at ScienceDirect

Preventive Medicine Reports

journal homepage: www.elsevier.com/locate/pmedr

Short communication

Health symptoms in residents living near shale gas activity: A retrospective record review from the Environmental Health Project

Beth Weinberger[a,*], Lydia H. Greiner[b], Leslie Walleigh[c], David Brown[a]

[a] Southwest Pennsylvania Environmental Health Project, The Grove, 760 Chapel St., New Haven, CT 06510, United States
[b] Southwest Pennsylvania Environmental Health Project, 2001 Waterdam Plaza Suite 201, McMurray, PA 15317, United States
[c] Southwest Pennsylvania Environmental Health Project, 2001 Waterdam Plaza Suite 201, McMurray, PA 15317, United States

ARTICLE INFO

Keywords:
Hydraulic fracturing
Health
Fracking
Shale gas
Unconventional gas

ABSTRACT

Increasing evidence demonstrates an association between health symptoms and exposure to unconventional natural gas development (UNGD). The purpose of this study is to describe the health of adults in communities with intense UNGD who presented for evaluation of symptoms. Records of 135 structured health assessments conducted between February 2012 and October 2015 were reviewed retrospectively. Publicly available data were used to determine proximity to gas wells. Analysis was restricted to records of adults who lived within 1 km of a well in Pennsylvania and denied employment in the gas industry ($n = 51$). Symptoms in each record were reviewed by a physician. Symptoms that could be explained by pre-existing or concurrent conditions or social history and those that began or worsened prior to exposure were excluded. Exposure was calculated using date of well drilling within 1 km. The number of symptoms/participant ranged from 0 to 19 (mean = 6.2; SD = 5.1). Symptoms most commonly reported were: sleep disruption, headache, throat irritation, stress or anxiety, cough, shortness of breath, sinus problems, fatigue, nausea, and wheezing. These results are consistent with findings of prior studies using self-report without physician review. In comparison, our results are strengthened by the collection of health data by a health care provider, critical review of symptoms for possible alternative causes, and confirmation of timing of exposure to unconventional natural gas well relative to symptom onset or exacerbation. Our findings confirm earlier studies and add to the growing body of evidence of the association between symptoms and exposure to UNGD.

1. Background

The public's health should be a consideration when there is widespread adoption of new industrial activity such as extraction of natural gas through hydraulic fracturing, commonly referred to as “fracking”. Hydraulic fracturing, the injection of pressurized water, chemicals and sand into a well bore to increase production of oil or gas, was first used in conventional vertical wells drilled into discrete oil or gas reservoirs. In recent years, the development of high volume, high pressure hydraulic fracturing, combined with directional drilling, has facilitated the extraction of oil and gas from unconventional reservoirs, such as shale and other “tight” geologic formations, where the oil and gas is distributed throughout the formation rather than in defined reservoirs. Proponents of hydraulic fracturing cite benefits such as reduced dependence on foreign oil and job creation in local communities. Public health professionals and others have raised concerns about short- and long-term health and environmental impacts.

Hydraulic fracturing is part of a larger process of extracting, processing and transporting natural gas. Taken together, it is referred to as unconventional natural gas development (UNGD). UNGD sites include well pads, where the hydraulic fracturing occurs, compressor stations, metering stations, and processing plants, all of which release emissions.

Air and water monitoring near well pads have documented the presence of multiple compounds with known human health effects, both short- and long-term. Compounds of concern are volatile organic compounds including benzene, associated with short-term effects of headache and dizziness and long-term effects of aplastic anemia and leukemia (ATSDR, 2015); toluene, associated with headaches, sleepiness, confusion, and possible permanent neurological damage (ATSDR, 2011a) ethylbenzene, associated with symptoms of eye and throat irritation and a possible carcinogen (ATSDR, 2011b) and xylene, associated with eye, nose, throat, and skin irritation and possible long-term neurologic effects (CCOHS, 2017).

Other compounds with documented adverse health outcomes

* Corresponding author.
E-mail addresses: bweinberger@environmentalhealthproject.org (B. Weinberger), lgreiner@environmentalhealthproject.org (L.H. Greiner), lwalleigh@environmentalhealthproject.org (L. Walleigh), bweinberger@environmentalhealthproject.org (D. Brown).

http://dx.doi.org/10.1016/j.pmedr.2017.09.002
Received 24 March 2017; Received in revised form 11 July 2017; Accepted 1 September 2017
Available online 14 September 2017

include particulate matter, associated with asthma attacks, acute bronchitis, and reduced lung function (OSHA, 2013), methylene chloride, associated with cancer (ATSDR, 2011c), and hydrogen sulfide, associated with eye, nose, and throat irritation and asthma (ATSDR, 2011d). Our understanding of the human health impacts of exposure, however, is hampered by the absence of human toxicity information on 75–80% of the chemicals used in this process (Elliott et al., 2016). In addition to chemical emissions, UNGD produces noise and light exposures at levels that may increase the risk of adverse health outcomes, including annoyance, sleep disturbance, and cardiovascular symptoms (Hays et al., 2017).

Self-report studies have consistently documented skin irritation and rash; respiratory symptoms including difficulty breathing; nose, throat, and sinus problems; gastrointestinal disturbances; headache; sleep disruption; and psychological symptoms including stress (Saberi, 2013; Ferrar et al., 2013; Rabinowitz et al., 2015; Steinzor et al., 2013). These studies relied on self-report of symptoms, obtained either through a survey "check-list" that was self-administered (Saberi, 2013; Steinzor et al., 2013) or administered by a research assistant (Rabinowitz et al., 2015). In one study a semi-structured interview was used (Ferrar et al., 2013). With the exception of the study conducted by Rabinowitz and colleagues (Rabinowitz et al., 2015), these studies used convenience samples that ranged in size from 33 to 108. Rabinowitz et al. used randomized subject selection and did not refer explicitly to UNGD in the survey process. Two studies included an estimate of exposure. Steinzor et al. demonstrated compounds with known human health effects in air and water samples; symptoms reported by participants were consistent with these effects. Rabinowitz et al. found increased prevalence of skin and respiratory symptoms was associated with increased proximity to natural gas wells.

Limitations of the self-report studies include the use of convenience samples and possible recall bias on the part of the participant. Onset and/or exacerbation of self-reported symptoms may be subject to recall bias on the part of the participant, particularly if the participants have a high level of awareness of the risks associated with exposure and/or understand the purpose of the study. None of the self-report studies incorporated review of data by a health care provider.

More recently, several population-based studies using publicly available or health system data have documented an association with poor birth outcomes (Casey et al., 2015; McKenzie et al., 2014; Stacy et al., 2015) asthma exacerbation (Rasmussen et al., 2016), infant mortality (Busby and Mangano, 2017), and childhood acute lymphocytic leukemia.(McKenzie et al., 2017) One other study demonstrated an association with migraine, chronic rhinosinusitis, and fatigue, symptoms previously documented in the other self-report studies. (Tustin et al., 2016)

The purpose of the present study is to describe the symptoms reported in a sample of Pennsylvania residents who lived in close proximity to unconventional gas wells. We conducted a retrospective review of 135 health assessment records of individuals who live in the Marcellus Shale region of the United States. The health assessments had been conducted by family nurse practitioners in collaboration with an occupational medicine physician. Because available evidence suggests that health impacts are related to proximity to wells, with symptoms more likely in individuals who live in closer proximity to gas wells (Rabinowitz et al., 2015; Casey et al., 2015; McKenzie et al., 2014; Stacy et al., 2015; Rasmussen et al., 2016; McKenzie et al., 2017; Tustin et al., 2016), this review was restricted to the records of individuals who lived within 1 km of at least one gas well. The study was reviewed and approved by the Duquesne University Institutional Review Board.

2. Method

Family nurse practitioners at the Southwest Pennsylvania Environmental Health Project (EHP) have been systematically collecting health data from residents of communities located near UNGD sites since 2012. This service was developed to meet the needs of residents who were concerned about health impacts and who sought evaluation by a health care professional. Services are advertised on the EHP website, local media, community meetings, and word-of-mouth and are offered at no charge. The health records of these clients provide a dataset of health symptoms reported by those living in proximity to UNGD sites.

Between February 1, 2012 and October 31, 2015, 135 children and adults completed the standardized health assessment, typically conducted face-to-face by a family nurse practitioner. The health assessments were conducted according to standard clinical practice for collecting a medical history and included current problems, review of systems, past medical history, family history, and social history. When indicated by the interview, a targeted physical examination was conducted. Individuals who completed this health assessment did so for their own personal health information.

All 135 records were reviewed by a team of health care providers that included a physician who is board certified occupational medicine (LW) and at least one nurse practitioner. Records were excluded if they were incomplete at the time of the review ($n = 2$); the client was < 18 years of age ($n = 21$); the client reported employment in the gas industry ($n = 7$); client resided in a state other than Pennsylvania ($n = 28$); client did not report any symptoms at the time of the health assessment ($n = 3$). After these exclusion criteria were applied, 74 records remained.

2.1. Proximity to unconventional natural gas wells

One author (BW) used publicly available data to determine the number of unconventional natural gas wells located within 1 km of each residence for the 74 records. Publicly available data includes location and "SPUD" date, or date drilling began. Using ArcGIS, the home address was used to calculate the distance from the home to the nearest well(s). Records were excluded if it was not possible to verify at least one gas well within 1 km of the residence ($n = 23$). After this criterion was applied, 51 records remained.

2.2. Symptom inclusion criteria

Prior to review of the records, the physician (LW) and nurse practitioner developed and implemented the symptom inclusion criteria. Each symptom recorded in the health assessment was reviewed in the context of past medical and surgical history, concurrent medical conditions, family and social history, and environmental exposures unrelated to UNGD. If a plausible cause for the symptom was identified, the symptom was not included in the analysis. For example, if the social history indicated a ½ pack/day smoking history, the symptom of "difficulty breathing" was not included. Symptoms were included only when there was no possible cause evident in the health assessment record. The records were not reviewed with the intent of establishing or confirming a diagnosis, but to determine if a plausible explanation for the symptom could be identified.

Independently, BW determined timing of the exposure for each symptom that met the inclusion criteria, using the SPUD date for each unconventional natural gas well within 1 km. The earliest SPUD date for wells within 1 km of the residence was considered the beginning of exposure to UNGD. The date of onset/exacerbation of each symptom was available in the health assessment record. If the date of onset/exacerbation of a symptom occurred prior to the earliest SPUD date for wells within 1 km, that symptom was not included in the analysis. Symptoms were included only if the onset/exacerbation occurred after the date of first exposure, estimated by the earliest SPUD date.

Descriptive statistics were used to determine frequency, distribution, and variance.

Table 1
Symptoms meeting inclusion criteria that were reported between February 2012 and October 2015 by 51 adults who lived within 1 km of an unconventional natural gas well in Pennsylvania.

Symptoms	# Reporting	% Reporting
Sleep disruption	22	43.1%
Headache	21	41.2%
Throat irritation	20	39.2%
Stress/anxiety	19	37.3%
Cough	17	33.3%
Shortness of breath	15	29.4%
Sinus problems	15	29.4%
Fatigue	12	23.5%
Nausea	12	23.5%
Wheezing	11	21.6%
Itchy eyes	11	21.6%
Weak/drowsy	9	17.6%
Abdominal pain	9	17.6%
Irritable moody	9	17.6%
Painful/dry eyes	8	15.7%
Painful joints	8	15.7%
Rash	8	15.7%
Dizziness	8	15.7%
Nose bleeds	7	13.7%
Tinnitus	7	13.7%
Aches	7	13.7%
Memory - short term	7	13.7%
Numbness	7	13.7%
Chest pain	6	11.8%
Hair loss	6	11.8%
Itchy skin	6	11.8%
Worry	6	11.8%
Palpitation	5	9.8%
Skin lesions/blisters	5	9.8%

3. Results

The 51 adults included in this record review had reported at least one symptom on their health assessment, denied occupation exposure related to natural gas extraction and lived in Pennsylvania within 1 km of an unconventional natural gas well. The average age of this sample was 57 (SD = 12.3), with a range of 24–85. More than half (56.8%) were female and the majority (83%) were married. Each individual lived within 1 km of a gas well; the number of wells ranged from 1 to 16, (mean 5.7, SD 3.6). A total of three counties in Pennsylvania are represented in this sample: Washington (n = 47), Butler (n = 3), and Bedford (n = 1) counties.

In this sample, all individuals reported at least one symptom at the time of the health assessment. The number of symptoms reported ranged from 1 to 19, with an average of 7.2 (SD = 4.9). Not all of the symptoms reported met the inclusion criteria (i.e., symptoms began or worsened after exposure to UNGD and could not be explained by a pre-existing or concurrent health condition). Some symptoms reported by 19 individuals (37%) did not meet inclusion criteria and were excluded, although the individuals remained in the analysis. The number of symptoms excluded/individual ranged from 1 to 7, with an average of 2.4 symptoms. For five of the 19 individuals, all reported symptoms were excluded.

The number of symptoms meeting inclusion criteria ranged from 0 to 19 with a mean of 6.2 (SD = 5.1) symptoms/individual. The most frequently reported symptoms that met inclusion criteria were sleep disturbance, headache, throat irritation, stress/anxiety, cough, shortness of breath, sinus, fatigue, wheezing, nausea (> 20% of sample).

Symptoms shown in Table 1 were reported by at least 10% of the sample. Symptoms not shown on Table 1, reported by < 10% of the sample were: weight change, hearing loss, vomiting, burning skin, and depression.

4. Discussion

The symptoms reported by residents of southwestern Pennsylvania who live within 1 km of an unconventional natural gas well are consistent with those reported in other self-report studies. The most commonly reported symptoms in this sample of adults were sleep disruption, headache, throat irritation, stress/anxiety, cough, shortness of breath, sinus problems, fatigue, nausea, and wheezing.

Limitations of this study include use of self-report data and a convenience sample. However, our methodology mitigates some of the limitations typically associated with this type of data and strengthens our results. Reported symptoms were abstracted from health records obtained by a nurse practitioner in consultation with a physician. Each symptom was evaluated using criteria to establish onset or exacerbation of the symptom relative to exposure to UNGD and to rule out other plausible explanations for the symptom. Only those symptoms that could not be explained by evidence in the health record (i.e., medical, surgical, or social history) and had a date of onset or exacerbation after exposure to UNGD began were included in the analysis.

Both the collection of symptom data, and the inclusion criteria used, distinguish this study from others that rely only on self-report. In comparison to such studies, our results are strengthened by the collection of health assessment data by a health care provider, critical review of symptoms for possible alternative causes, and confirmation of timing of exposure relative to symptom onset or exacerbation.

Health care providers whose clients live or work in communities where unconventional techniques are used to extract natural gas and/or oil should be alert to the possibility of environmental exposures. Symptoms, particularly those that are unexplained by concurrent medical conditions, may be related to environmental exposures.

Funding

This work was supported by Heinz Endowments (E4442), Pittsburgh, PA. Heinz Endowments had no role in study design; collection, analysis and interpretation of data; writing of the report; or the decision to submit the article for publication.

Conflict of interest

The authors declare no conflict of interest.

References

ATSDR, 2011a. Toluene. Available at: Agency for Toxic Substances and Disease Registry Toxic Substances Portalhttp://www.atsdr.cdc.gov/substances/toxsubstance.asp?toxid=29, Accessed date: 13 February 2017.
ATSDR, 2011b. Ethylbenzene. Available at: Agency for Toxic Substances and Disease Registry Toxic Substances Portalhttps://www.atsdr.cdc.gov/substances/toxsubstance.asp?toxid=66, Accessed date: 13 February 2017.
ATSDR, 2011c. Methylene chloride. Available at: Agency for Toxic Substances and Disease Registry Toxic Substances Portalhttp://www.atsdr.cdc.gov/substances/toxsubstance.asp?toxid=42, Accessed date: 13 February 2017.
ATSDR, 2011d. Hydrogen sulfide carbonyl sulfide. Available at: Agency for Toxic Substances and Disease Registry Toxic Substances Portalhttp://www.atsdr.cdc.gov/substances/toxsubstance.asp?toxid=67, Accessed date: 13 February 2017.
ATSDR, 2015. ToxFAQsTM for benzene. Available at: Agency for Toxic Substances and Disease Registry Toxic Substances Portalhttps://www.atsdr.cdc.gov/toxfaqs/tf.asp?id=38&tid=14, Accessed date: 13 February 2017.
Busby, C., Mangano, J., 2017. There's a world going on underground—infant mortality and fracking in Pennsylvania. J. Environ. Prot. 8, 381–393.
Casey, J.A., et al., 2015. Unconventional natural gas development and birth outcomes in Pennsylvania, USA. Epidimiology. http://dx.doi.org/10.1097/EDE.0000000000000387.
CCOHS, 2017. Xylene. Available at: Canadian Centre for Occupational Health and Safety OSH Answers Fact Sheethttp://www.ccohs.ca/oshanswers/chemicals/chem_profiles/xylene.html, Accessed date: 13 February 2017.
Elliott, E.G., Ettinger, A.S., Leaderer, B.P., Bracken, M.B., Deziel, N.C., 2016. A systematic evaluation of chemicals in hydraulic-fracturing fluids and wastewater for reproductive and developmental toxicity. J. Expo. Sci. Environ. Epidemiol. http://dx.doi.org/10.1038/jes.2015.81.
Ferrar, K.J., et al., 2013. Assessment and longitudinal analysis of health impacts and

stressors perceived to result from unconventional shale gas development in the Marcellus Shale region. Int. J. Occup. Environ. Health 19, 104–112.
Hays, J., McCawley, M., Shonkoff, S.B.C., 2017. Public health implications of environmental noise associated with unconventional oil and gas development. Sci. Total Environ. 580, 448–456.
McKenzie, L.M., et al., 2014. Birth outcomes and maternal residential proximity to natural gas development in rural Colorado. Environ. Health Perspect. 122, 412–417.
McKenzie, L.M., et al., 2017. Childhood hematologic cancer and residential proximity to oil and gas development. PLoS One 12, e0170423.
OSHA, 2013. Diesel exhaust. Available at: Occupational Health and Safety Administration Safety and Health Topicshttps://www.osha.gov/SLTC/dieselexhaust/.
Rabinowitz, P.M., et al., 2015. Proximity to natural gas wells and reported health status: results of a household survey in Washington County, Pennsylvania. Environ. Health Perspect. 123, 21–26.
Rasmussen, S.G., et al., 2016. Association between unconventional natural gas development in the Marcellus Shale and asthma exacerbations. JAMA Intern. Med. http://dx.doi.org/10.1001/jamainternmed.2016.2436.
Saberi, P., 2013. Navigating medical issues in shale territory. New Solut. J. Environ. Occup. Health Policy 23, 209–221.
Stacy, S.L., et al., 2015. Perinatal outcomes and unconventional natural gas operations in Southwest Pennsylvania. PLoS One 10, e0126425.
Steinzor, N., Subra, W., Sumi, L., 2013. Investigating links between shale gas development and health impacts through a community survey project in Pennsylvania. New Solut. J. Environ. Occup. Health Policy 23, 55–83.
Tustin, A.W., et al., 2016. Associations between unconventional natural gas development and nasal and sinus, migraine headache, and fatigue symptoms in Pennsylvania. Environ. Health Perspect. http://dx.doi.org/10.1289/EHP281.

ATTACHMENT C

STUDY 13

IEA
International Epidemiological Association

International Journal of Epidemiology, 2022, 525–536
https://doi.org/10.1093/ije/dyab246
Advance Access Publication Date: 13 December 2021
Original article

OXFORD

Environmental Epidemiology

Associations between residential proximity to oil and gas extraction and hypertensive conditions during pregnancy: a difference-in-differences analysis in Texas, 1996–2009

Mary D Willis ,[1,2]* **Elaine L Hill,**[3] **Molly L Kile,**[1] **Susan Carozza**[1] **and Perry Hystad**[1]

[1]School of Biological and Population Health Sciences, College of Public Health and Human Sciences, Oregon State University, Corvallis, OR, USA, [2]Department of Epidemiology, School of Public Health, Boston University, Boston, MA, USA and [3]Department of Public Health Sciences, School of Medicine and Dentistry, University of Rochester, Rochester, NY, USA

*Corresponding author. 160 SW 26th St, Corvallis, OR 97331, USA. E-mail: mary.willis@oregonstate.edu

Received 1 October 2020; Editorial decision 29 September 2021; Accepted 9 November 2021

Abstract

Background: Oil and gas extraction produces air pollutants that are associated with increased risks of hypertension. To date, no study has examined residential proximity to oil and gas extraction and hypertensive conditions during pregnancy. This study quantifies associations between residential proximity to oil and gas development on gestational hypertension and eclampsia.

Methods: We utilized a population-based retrospective birth cohort in Texas (1996–2009), where mothers reside <10 km from an active or future drilling site ($n = 2\,845\,144$.) Using full-address data, we linked each maternal residence at delivery to assign exposure and evaluate this exposure with respect to gestational hypertension and eclampsia. In a difference-in-differences framework, we model the interaction between maternal health before (unexposed) or after (exposed) the start of drilling activity (exposed) and residential proximity near (0–1, >1–2 or >2–3 km) or far (≥3–10 km) from an active or future drilling site.

Results: Among pregnant women residing 0–1 km from an active oil or gas extraction site, we estimate 5% increased odds of gestational hypertension [95% confidence interval (CI): 1.00, 1.10] and 26% increased odds of eclampsia (95% CI: 1.05, 1.51) in adjusted models. This association dissipates in the 1- to 3-km buffer zones. In restricted models, we find elevated odds ratios among maternal ages ≤35 years at delivery, maternal non-Hispanic White race, ≥30 lbs gained during pregnancy, nulliparous mothers and maternal educational attainment beyond high school.

Conclusions: Living within 1 km of an oil or gas extraction site during pregnancy is associated with increased odds of hypertensive conditions during pregnancy.

Key words: Gestational hypertension, eclampsia, oil drilling, gas drilling, resource extraction, difference-in-differences

Key Messages

- Among pregnant women who reside within 1 km of at least one oil or gas drilling site, we find 5% increased odds of gestational hypertension and 26% increased odds of eclampsia.
- We find no evidence of this association for pregnant women who reside in within 1–2 km or 2–3 km of at least one oil or gas drilling site.
- Restricted models show that women who were most sensitive to drilling exposures were under age 35 years, were nulliparous, were non-Hispanic White mothers and had greater than a high-school education.
- However, we also find an unexpected reduced association among Hispanic women and women with less than a high-school education.
- Given that hypertensive conditions during pregnancy carry serious risks for pregnant women and their infants, these findings warrant further examination.

Background

Gestational hypertension, pre-eclampsia and eclampsia are hypertensive conditions in pregnancy that threaten maternal health. Up to 8% of all pregnancies are impacted by hypertensive conditions and 16% of maternal deaths are attributed to complications arising from high blood pressure.[1] Gestational hypertension, which is defined as incident blood pressure of >140/90 at two time points after 20 weeks of pregnancy, is the most common condition. Pre-eclampsia and eclampsia, though less common, are more serious complications.[2] These hypertensive conditions have significant public health consequences and cost the healthcare system over 1 billion dollars for mothers within 12 months of delivery.[3]

Over the past 15 years, incidence of hypertensive conditions during pregnancy have increased.[4,5] Known risk factors such as obesity, nulliparity and history of hypertension explain some of these cases,[4–6] but the complex mechanisms causing this increase are largely unknown.[7] Environmental contamination, particularly air pollution, is emerging as a contributor towards high blood pressure during pregnancy.[8–12] A recent expert review by the National Toxicology Program concludes that components of traffic-related air pollution may be causally linked to gestational hypertension, pre-eclampsia and eclampsia.[13] However, few other sources of air pollution have been assessed with respect to maternal hypertensive conditions.

We hypothesized that air-pollution exposures from oil and gas resource extraction may pose risks for hypertensive conditions during pregnancy similarly to traffic-related air pollution. Additional sources of pollution released from the oil and gas industry such as water contamination and increased noise and light pollution may also negatively impact maternal-health conditions.[14–17] Environmental monitoring studies show higher concentrations of air pollution and water contamination near oil and gas development sites compared with background levels[15,18–21] and a recent pair of biomonitoring studies shows differences in exposures to heavy metals and volatile organic compounds among pregnant women residing in close proximity to gas extraction compared with people living in the general population.[22,23] Therefore, we hypothesize that there may be an increased odds of gestational hypertension and eclampsia associated with increasing residential proximity to oil and gas drilling sites.

Globally, estimates indicate that 300 million people across six continents reside on viable oil and gas reservoirs that may be drilled in the future.[24] The oil and gas industry has rapidly expanded over the last 20 years, with an estimated 17.6 million Americans now living within 1.6 km (1 mile) of an active oil or gas drilling site.[25] Within this industry, there are many potential sources of air pollution including site construction, borehole drilling, hydraulic fracturing, industry traffic and gas flaring.[15] Air pollution from oil and gas drilling has some similar components to traffic-related air pollution such as particulate matter [diesel particulate matter (PM), PM_{10} (particulate matter <10 micrometers)], volatile organic compounds (benzene, toluene, ethylbenzene and xylene) and polycyclic aromatic hydrocarbons (naphthalene, chlorobenzene, phenol).[15] These pollutants are expected to be concentrated within 1 km of drilling sites and mostly dissipating to background levels beyond 3 km.[25]

Whereas oil and gas extraction increases air pollution, this industry may also produce positive community impacts, including increased employment and income, enhanced community resources and reduced oil and gas costs.[26–29] New employment prospects may also rapidly shift the socio-demographic composition of a community.[30,31] These socio-demographic and economic changes provide an important, but often overlooked, source of confounding that needs to be taken into account to determine the population health impacts of pollution from this industry. Despite extensive community concerns about the public health implications of oil and gas development, the population health implications of this industry are unclear. Residing near oil and gas development has been associated with a number of conditions such as asthma exacerbations,[32–34] anxiety or depression,[35,36] sleep disturbances[37] and adverse birth outcomes.[38–47] A pilot study that examined markers of cardiovascular disease observed increased systolic blood pressure for participants who lived closer to drilling sites than those who lived farther away.[48] To date, no studies have specifically examined associations between oil and gas development and hypertensive conditions during pregnancy.

Locations of oil and gas drilling sites and who chooses to live near them is not entirely random, so conventional epidemiological approaches may not adequately account for the socio-economic and structural factors that lead to living near an oil and gas site, independently of pollution effects. To overcome this problem, we implement a difference-in-differences analysis to evaluate associations between drilling exposures and hypertensive conditions during pregnancy. This technique allows us to compare before and after drilling changes in hypertension outcome risk to a nearby temporal control group where drilling has not directly occurred[49] and to disaggregate the socio-economic impacts of an industrial boom from the simultaneous introduction of new environmental pollution. In this study design, we aim to overcome the confounding that may be induced from changes in maternal stress or anxiety during pregnancy,[35,50] variation in healthcare usage[51,52] or other external non-environmental factors that may be related to an industrial boom. Using geocoded vital-statistics records from 1996 to 2009 in Texas, the state with the highest oil and gas production during a period of rapid industry growth,[53] we conduct a population-based retrospective cohort study in a difference-in-differences analytical framework to assess associations between exposure to oil and gas drilling and hypertensive conditions during pregnancy.

Methods

Data sources

This cohort study evaluates birth-certificate data obtained from the Vital Statistics Program in the Texas Department of State Health Services for the period of 1 January 1996 to 31 December 2009. The restricted-access data contain maternal residential location at delivery geocoded to the full-address level. We received academic access to a propriety database of oil and gas drilling sites from Enverus Drillinginfo.[54] This study has been approved by the Institutional Review Boards at Oregon State University (#6692) and the Texas Department of State Health Services (#15–063). We used the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guidelines.[55]

Study population

We acquired birth-certificate data for all births in Texas from 1 January 1996 to 31 December 2009, which contain the residences at delivery geocoded to the full-address level ($N = 4\,569\,428$). We exclude non-singleton births ($n = 131\,880$) and remove implausible observations ($n = 12\,577$) based on maternal age (≤ 10 and ≥ 65 years old), gestational age (<22 and ≥ 45 weeks) and birthweight (≤ 500 and ≥ 5000 g). In addition, we remove observations with missing continuous covariates ($n = 414\,816$), which is primarily due to the weight-gain-during-pregnancy variable. To reduce potential for community-level confounding, our study population further excluded mothers living outside 10 km of an active or permitted drilling site between 1 January 1985 and 30 June 2019 ($n = 1\,165\,011$). Our study population contains mothers who gave birth at 22–44 weeks' gestation with a reported residence at delivery within 10 km of an active or future drilling site ($n = 2\,845\,144$ mothers.)

Exposure assessment

We evaluate exposure to oil and gas extraction via maternal residential proximity at delivery to at least one active drilling site on the date of delivery. All drilling sites with oil or gas as the primary resource with a first date of drilling of between 1 January 1985 and 30 June 2019 were included in our database. Including drilling-site activity prior to our study period that were active during pregnancy allows for better understanding of the full extent of oil and gas extraction in a community, whereas drilling activity after our study period provides a reasonable counterfactual for places where oil and gas extraction will eventually occur but has not started yet. Any drilling site with an end date for its activity prior to the delivery date was excluded from this exposure metric.

A recent review on environmental exposures from oil and gas activity concludes that pollution directly from drilling should dissipate to background levels at 3 km from the drilling site, with the highest concentrations within 1 km.[25] To examine potential exposure–response gradients, we split our sample into four distinct zones: 0 to <1, 1 to <2, 2 to <3 and 3–10 km. The 3- to 10-km group represents the sample that is unlikely to be exposed to air pollution from oil and gas drilling, thus the participants in this zone

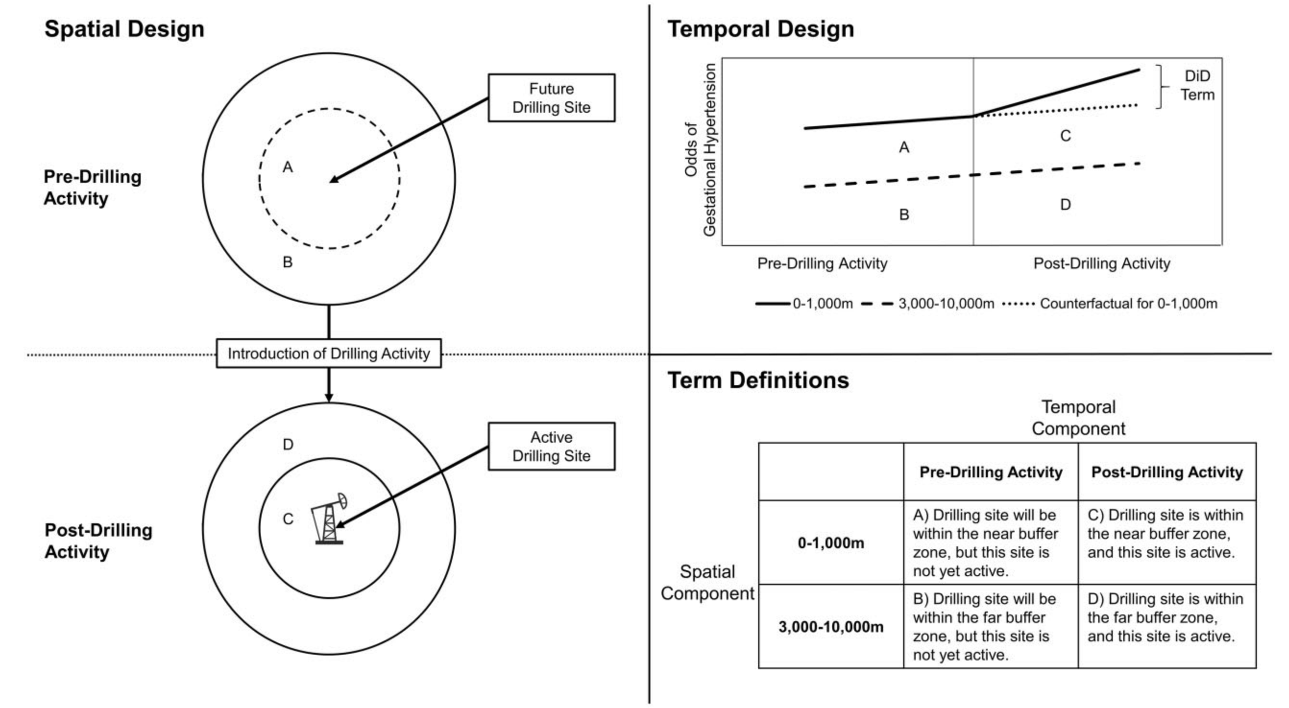


Figure 1 Conceptual diagram of spatial and temporal components for the difference-in-differences study design to examine the association between residential proximity to oil and gas drilling activity and hypertensive disorders during pregnancy. DiD, difference-in-differences.

act as the counterfactual in the difference-in-differences framework.

Outcome assessment

Data in birth certificates were used to assign a hypertension condition during pregnancy. This certificate has checkboxes for risk factors during pregnancy including gestational hypertension and eclampsia (yes if indicated, otherwise no). There was no question specific to pre-eclampsia for the study period, thus we were unable to examine this outcome. We treated both outcomes as a binary (reported diagnosis vs no reported diagnosis) in all models.

Statistical analysis

We compared trends in population demographics and socio-economic status before and after drilling began in our near and far groups using means in each subset of our sample. Because our analysis was predicated on drilling-related pollution as the potential cause of hypertension, we then examined trends in maternal hypertensive conditions before drilling began in our near and far groups. We also examined distance gradients in risk by implementing unadjusted local regressions of distance to the nearest active drilling site and our outcomes.

We subsequently implemented logistic-regression models with robust standard errors on our sample using a difference-in-differences framework in which the coefficient of interest was the interaction between an indicator for residential location within a buffer (0–1, 1–2 and 2–3 km) of one or more drilling sites and an indicator for delivery date after drilling began (Figure 1). In this model, we compared pregnant people who reside near drilling before and after drilling begins to a temporal control group in similar communities.[49] By leveraging a difference-in-differences design, we were able to parse the impacts of drilling-related pollution on maternal health from secular trends in hypertension treatment that may be changing over time. This study design takes into account unmeasured confounding factors that may influence the magnitude of our associations. Minimally adjusted models are adjusted for birth year (categorical). Fully adjusted models contained a priori selected covariates via literature review that were potential confounders or risk factors: birth year (categorical), infant sex (male, female), gestational age (continuous), maternal age (continuous), maternal race and ethnicity (White non-Hispanic, Black non-Hispanic, Hispanic, other/unknown/missing),[56] maternal educational attainment (less than high school, high-school graduate, some college education, bachelor's degree, postgraduate education, missing), nulliparous (yes, no), prenatal care received (yes, no, missing), smoking during pregnancy (yes, no, missing), maternal weight gain during pregnancy (continuous) and distance to major roads (continuous in metres). We implemented separate models for each buffer zone of the residences. To test the sensitivity of our

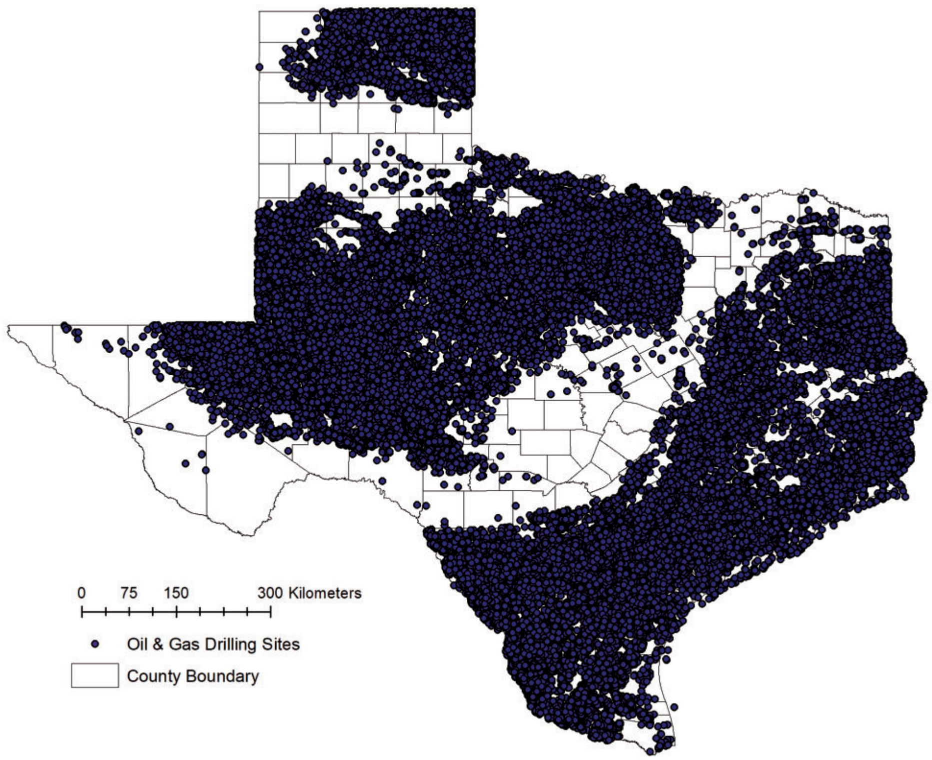


Figure 2 Spatial distribution of oil and gas drilling in Texas, 1985–2019. Data displayed include all oil and gas drilling sites spudded in Texas between 1 January 1985 and 30 June 2019.

results, we present several variations on our adjusted model as follows: (i) adding neighbourhood covariates at the census tract level (unemployment percentage, White-population percentage, median household income in US dollars); (ii) adding a month covariate to account for seasonal variability; (iii) removing the covariates related to race, ethnicity and educational attainment to check for confounding by socio-demographic variation; (iv) excluding any records with missing data indicators to examine the role of missing data (e.g. conduct a complete case analysis); and (v) excluding births in 2008–2009 to assess the impact of the Great Recession.

Risks of gestational hypertension and eclampsia vary by demographic characteristics, socio-economic conditions and pregnancy attributes.[4,7,57] We examined variations in associations by key risk factors in restricted models: maternal age ($\leq$35 years, >35 years), maternal weight gain during pregnancy (<30 lbs, $\geq$30 lbs), parity (nulliparous, multiparous), maternal race and ethnicity (White non-Hispanic, Black non-Hispanic, Hispanic) and maternal educational attainment (high-school diploma or less, more education than high-school diploma). These cut-off points were selected based on risk factors for both conditions, as well as the study characteristics in this sample. Due to concerns about incorporating an additional interaction term into our difference-in-differences models, we implement restricted models for each subgroup by buffer zone of the residences: 0–1, 1–2 and 2–3 km.

Results

The spatial distributions of drilling activity by resource and type across the state are displayed in Figure 2. After accounting for secular trends using the difference-in-differences framework, descriptive statistics showed that the percentage of women reporting a gestational hypertension diagnosis increases by 0.3% and the percentage of women reporting an eclampsia diagnosis decreases by −0.1% for women residing within 1 km vs women residing 3–10 km from at least one drilling site (Table 1). Maternal characteristics were largely similar in the difference-in-differences terms, except for proportions of Black non-Hispanic women (−7.8%) and Hispanic women (7.9%). Characteristics for the 1- to 2- and 2- to 3-km groups showed similar patterns (Supplementary Tables S1 and S2, available as Supplementary data at *IJE* online). Gestational hypertension and eclampsia diagnoses over time appear to be similar among our groups, though there is annual

Table 1 Demographic information for the Texas birth cohort (1996–2009) for maternal residences <1 and 3–10 km away before and after drilling began

Characteristic	Near, 0–1 km		Far, 3–10 km		Differences between groups		
	Pre-drilling activity A	Post-drilling activity B	Pre-drilling activity C	Post-drilling activity D	B–A[a]	D–C[a]	DiD term[b]
Total births	86 893	158 644	438 370	1 746 922	—	—	—
Gestational hypertension (%)	4.2	4.4	4.2	4.1	0.2	−0.1	0.3
Eclampsia (%)	0.3	0.2	0.2	0.2	−0.1	0.0	−0.1
Female sex (%)	49.1	48.9	48.9	48.9	−0.2	0.0	−0.2
Gestational age (mean)	38.7	38.6	38.7	38.6	−0.1	−0.1	0.0
Maternal age (mean)	26.1	26.0	26.2	26.1	−0.1	−0.1	0.0
Maternal race and ethnicity							
White non-Hispanic (%)	46.1	47.1	35.9	35.3	1.0	−0.6	1.6
Black non-Hispanic (%)	14.9	9.7	9.8	12.4	−5.2	2.6	−7.8
Hispanic (%)	34.2	40.2	50.2	48.3	6.0	−1.9	7.9
Other (%)	4.7	3.0	4.3	4.1	−1.7	−0.2	−1.5
Maternal educational attainment							
Did not complete high school (%)	35.5	28.3	34.3	31.7	−7.2	−2.6	−4.6
Completed high school (%)	32.7	30.3	28.3	29.9	−2.4	1.6	−4.0
Some college (%)	18.7	23.6	17.5	20.3	4.9	2.8	2.1
Bachelor's degree (%)	11.4	12.2	11.4	11.0	0.8	−0.4	1.2
Postgraduate (%)	6.1	5.2	7.3	6.4	−0.9	−0.9	0.0
Weight gain during pregnancy (lbs)	30.4	31.0	29.9	30.2	0.6	0.3	0.3
Nulliparous (%)	41.1	40.0	41.7	40.3	−1.1	−1.4	0.3
No prenatal care (%)	1.6	2.0	1.9	2.2	0.4	0.3	0.1
Smoking during pregnancy	8.4	10.3	6.6	7.6	1.9	1.0	0.9
Neighbourhood characteristics[c]							
Nearest highway (m)[d]	1482	1700	1247	1501	218	254	−36
Median household income (USD)	44 452	47 282	41 258	44 004	2830	2746	84
Unemployment (%)	4.0	6.1	5.3	6.0	2.1	0.7	1.4
White population (%)	64.9	72.8	66.5	66.3	7.9	−0.2	8.1

DiD, difference-in-differences.

[a]A vs B and C vs D columns are the result of the differences for binary and continuous characteristics to compare demographic characteristics before and after drilling began, where categorical covariates are reassigned as dummy indicator variables.

[b]The DiD term is the difference of the near (0–1 km) before and after to the far (3–10 km) groups between near and exposed on the demographic characteristic.

[c]Derived from the US Census at the tract level. Births before 2005 were joined to the 2000 Census data and births in 2005 and after were joined to the 2010 Census data.

[d]Derived from the 2010 Census road file for Texas.

variation (Supplementary Figure S1, available as Supplementary data at *IJE* online).

Minimally adjusted and fully adjusted results show largely similar results (Table 2). Our fully adjusted difference-in-differences models for all women in our sample showed an increased odds of reporting gestational hypertension [1.05; 95% confidence interval (CI): 1.00, 1.09) and eclampsia (1.26; 95% CI: 1.05, 1.51) among women who resided within 1 km of at least one active drilling site at delivery compared with other women living within 1 km of a drilling site before active drilling and women living within 3–10 km of a drilling site before or after active drilling. These effects dissipated for gestational hypertension at 1–2 km (0.99; 95% CI: 0.95, 1.04) and 2–3 km (1.00; 95% CI: 0.94, 1.05). Although the point estimates were still elevated at 1–2 km for eclampsia, the association becomes statistically null at 1–2 km (1.10; 95% CI: 0.92, 1.32) and 2–3 km (0.97; 95% CI: 0.78, 1.21).

We then proceeded to examine restricted models (Table 3). We found elevated odds of gestational hypertension at 1 km in restricted models among women <35 years old at delivery (1.05; 95% CI: 1.00, 1.10), weight gain during pregnancy of >30 lbs (1.08; 95% CI: 1.02, 1.15), nulliparous women (1.13; 95% CI: 1.06, 1.20), White non-Hispanic women (1.27; 95% CI: 1.19, 1.35) and women with more education than a high-school diploma

Table 2 Difference-in-differences estimates (95% confidence intervals) between maternal residential distances of at least one active drilling site and markers of hypertensive conditions during pregnancy by key risk factors

Minimally adjusted model	*n*	Gestational hypertension	Eclampsia
0–1 km	2 430 829	1.04 (0.99, 1.08)	1.25 (1.04, 1.49)
1–2 km	2 429 660	0.97 (0.93, 1.02)	1.16 (0.97, 1.39)
2–3 km	2 355 239	0.98 (0.93, 1.04)	1.03 (0.83, 1.28)
Fully adjusted model	*n*	Gestational hypertension	Eclampsia
0–1 km	2 430 829	1.05 (1.00, 1.10)	1.26 (1.05, 1.51)
1–2 km	2 429 660	0.99 (0.95, 1.04)	1.10 (0.92, 1.32)
2–3 km	2 355 239	1.00 (0.94, 1.05)	0.97 (0.78, 1.21)

Reported coefficient is the interaction term for residence in that distance bin of drilling and after drilling has started. Minimally adjusted model is a logistic regression with adjustment for birth year (categorical for each year from 1996 to 2009). Fully adjusted model is a logistic regression with adjustment for birth year (categorical), infant sex (male, female), gestational age (continuous), maternal age (continuous), maternal race and ethnicity (White non-Hispanic, Black non-Hispanic, Hispanic, other/unknown/missing), maternal educational attainment (less than high school, high-school graduate, some college education, bachelor's degree, postgraduate education, missing), nulliparous (yes, no), prenatal care received (yes, no, missing), smoking during pregnancy (yes, no, missing), maternal weight gain during pregnancy (continuous) and distance to major roads (continuous in metres). All models include robust standard errors.

(1.29; 95% CI: 1.20, 1.40) (Table 2.) These effects persisted to 1–2 km for nulliparous women, White non-Hispanic women and women with more education than a high-school diploma. Protective effects out to 2 km were also noted for Hispanic women and women with less than or equivalent to a high-school diploma.

We also observed evidence for elevated odds of eclampsia in restricted models among women <35 years old at delivery (1.25; 95% CI: 1.04, 1.51), nulliparous women (1.62; 95% CI: 1.28, 2.06) and women with more education than a high-school diploma (1.58; 95% CI: 1.13, 2.22) (Table 3). Restricted models for weight gain yielded odds ratios of similar magnitude for <30 lbs of weight gain (1.32; 95% CI: 1.01, 1.72) and >30 lbs of weight gain (1.21; 95% CI: 0.95, 1.55). These effects persisted out to 2 km for nulliparous women and weight gain during pregnancy of >30 lbs.

Sensitivity analyses of these results generally show consistent elevated point estimates, but many of these results contain less statistical precision than the primary adjusted model (Supplementary Table S3, available as Supplementary data at *IJE* online). The largest attenuation of model estimates occurred when neighbourhood covariates (census tract unemployment percentage, White-population percentage and median household income) were added to the model for gestational hypertension (1.03; 95% CI: 0.98, 1.07) and eclampsia (1.17; 95% CI: 0.97, 1.40). Adding a covariate for birth month yielded similar results to the primary models. Removing socio-demographic variables and observations with missing data from the model did not change model results, respectively. Removing birth years for 2008–2009 (corresponding to the Great Recession) increased model estimates for gestational hypertension (1.08; 95% CI: 1.03, 1.14) and eclampsia (1.34; 95% CI: 1.11, 1.61) among women who resided within 1 km of at least one active drilling site at delivery compared with other women living within 1 km of a drilling site before active drilling.

Discussion

This study represents the first analysis to examine the impacts of oil and gas drilling on hypertension conditions during pregnancy. By applying a difference-in-differences design, we attempted to disaggregate the socio-economic changes associated with an industrial boom from the impacts of residing near drilling-related pollution on maternal health. Specifically, our study found that women who resided within 1 km of at least one active drilling site at delivery had a 5% increased odds of reporting gestational hypertension and a 26% increased odds of reporting eclampsia. Women who were most sensitive to oil and gas drilling exposures were <35 years old, gained >30 lbs during pregnancy, were nulliparous, were non-Hispanic White mothers and had greater than a high-school education. These findings indicate that close residential proximity to oil and gas drilling may pose a substantial hazard for pregnant women.

Our results contribute to the growing body of literature on the population health impacts of the oil and gas extraction industry. Much of the focus to date has been on adverse birth outcomes, including preterm birth, birthweight, small for gestational age, congenital anomalies and infant mortality.[38–45] Existing analyses on drilling and infant health generally find elevated risks that persist much further than the 1 km that we see in our risk estimates for pregnancy-related hypertensive conditions. This smaller distance suggests that there may be distinct exposure pathways for drilling-related pollution to influence infant

Table 3 Difference-in-differences estimates (95% confidence intervals) between maternal residential distances of at least one active drilling site and markers of hypertensive conditions during pregnancy by key maternal characteristics

Maternal characteristic	*n*	Gestational hypertension	Eclampsia
Maternal age ≤35 years			
0–1 km	2 243 504	1.05 (1.00, 1.10)	1.25 (1.04, 1.51)
1–2 km	2 240 795	1.00 (0.95, 1.05)	1.14 (0.95, 1.38)
2–3 km	2 172 076	0.99 (0.94, 1.05)	0.96 (0.76, 1.20)
Maternal age >35 years			
0–1 km	187 325	1.01 (0.86, 1.18)	1.27 (0.65, 2.50)
1–2 km	188 819	0.96 (0.83, 1.12)	0.67 (0.35, 1.28)
2–3 km	183 122	1.02 (0.84, 1.22)	1.18 (0.50, 2.80)
Weight gain <30 lbs			
0–1 km	1 177 727	1.00 (0.93, 1.07)	1.32 (1.01, 1.72)
1–2 km	1 179 012	0.93 (0.87, 1.00)	0.86 (0.64, 1.14)
2–3 km	1 145 161	0.98 (0.90, 1.06)	1.03 (0.73, 1.45)
Weight gain ≥30 lbs			
0–1 km	1 253 102	1.08 (1.02, 1.15)	1.21 (0.95, 1.55)
1–2 km	1 250 648	1.04 (0.98, 1.10)	1.30 (1.02, 1.64)
2–3 km	1 210 078	1.01 (0.93, 1.08)	0.93 (0.70, 1.25)
Nulliparous			
0–1 km	986 253	1.13 (1.06, 1.20)	1.62 (1.28, 2.06)
1–2 km	985 917	1.07 (1.00, 1.14)	1.45 (1.14, 1.86)
2–3 km	955 430	0.99 (0.92, 1.07)	0.93 (0.70, 1.23)
Multiparous			
0–1 km	1 444 576	0.95 (0.89, 1.02)	0.87 (0.65, 1.15)
1–2 km	1 443 743	0.91 (0.85, 0.97)	0.74 (0.56, 0.98)
2–3 km	1 399 809	1.01 (0.93, 1.10)	1.03 (0.73, 1.47)
White non-Hispanic			
0–1 km	888 367	1.27 (1.19, 1.35)	1.18 (0.89, 1.55)
1–2 km	880 919	1.17 (1.10, 1.26)	1.26 (0.96, 1.67)
2–3 km	840 026	1.08 (0.99, 1.17)	1.16 (0.81, 1.65)
Black non-Hispanic			
0–1 km	287 093	0.90 (0.80, 1.01)	1.69 (1.06, 2.68)
1–2 km	287 801	0.86 (0.76, 0.98)	0.89 (0.61, 1.58)
2–3 km	278 214	0.87 (0.74, 1.03)	0.80 (0.43, 1.52)
Hispanic or Latina			
0–1 km	1 156 530	0.80 (0.74, 0.87)	1.13 (0.84, 1.52)
1–2 km	1 161 190	0.79 (0.73, 0.86)	0.96 (0.72, 1.28)
2–3 km	1 140 524	0.95 (0.86, 1.03)	0.86 (0.62, 1.18)
≤High school			
0–1 km	1 499 065	0.92 (0.86, 0.97)	1.14 (0.92, 1.42)
1–2 km	1 493 873	0.91 (0.85, 0.96)	1.04 (0.83, 1.30)
2–3 km	1 451 265	0.96 (0.89, 1.04)	1.04 (0.80, 1.36)
>High school			
0–1 km	913 573	1.29 (1.20, 1.40)	1.58 (1.13, 2.22)
1–2 km	917 335	1.11 (1.03, 1.20)	1.21 (0.89, 1.65)
2–3 km	885 749	1.04 (0.96, 1.14)	0.88 (0.59, 1.31)

Reported coefficient is the interaction term for residence in that distance bin of drilling and after drilling has started. Model is a logistic regression with adjustment for birth year (categorical), infant sex (male, female), gestational age (continuous), maternal age (continuous), maternal race and ethnicity (White non-Hispanic, Black non-Hispanic, Hispanic, other/unknown/missing), maternal educational attainment (less than high school, high-school graduate, some college education, bachelor's degree, postgraduate education, missing), nulliparous (yes, no), prenatal care received (yes, no, missing), smoking during pregnancy (yes, no, missing), maternal weight gain during pregnancy (continuous) and distance to major roads (continuous in metres). All models include robust standard errors.

health outcomes compared with maternal hypertension conditions. Components of pollution from oil and gas extraction sites are estimated to dissipate to background levels within 3 km of the extraction location, where the bulk of the dispersion is within 1 km for air pollution[25] and 2 km for water contamination.[18] This combination of evidence provides additional support for our results indicating that drilling activity is associated with hypertensive conditions during pregnancy.

We observed different socio-demographic characteristics of mothers that substantially altered the risk estimates for residing near at least one active drilling site and pregnancy-related hypertension. Surprisingly, these were opposite to the factors typically observed in environmental injustice cases in which higher exposure and risk are observed among minority and lower socio-economic-status populations.[58–61] In our analysis restricted to Hispanic women, we find a protective association between residential proximity to oil and gas drilling and odds of gestational hypertension. This result contrasts with some recent work on drilling-related exposures and population health outcomes.[34,45] Existing work has documented that Hispanic women show a decreased odds of a gestational hypertension diagnosis relative to White non-Hispanic women,[62] but we do not expect that this difference could explain our results. These protective results may be due to dissimilar residential patterns among Hispanic women living near oil and gas drilling sites or simply residual confounding in our analyses. Considering the severe health consequences of gestational hypertension and eclampsia, future research is necessary to fully understand which subpopulations may be disproportionately burdened by health impacts from exposure to oil and gas drilling.

This study has several strengths worth noting. First, we used a large population-based retrospective birth cohort to obtain our maternal-health information, yielding a sample size ($n = 2\,845\,144$) that is much larger than existing work on drilling activity and population health. This feature of our data allows us to examine smaller distances between drilling and residences without forfeiting the power to detect associations. Second, our study setting is Texas, the state with the most oil and gas production in the country,[53] and our study period covers multiple oil and gas booms.[63] Third, we apply a difference-in-differences analytical framework to our study population.[49] This feature of our study design reduces the potential for residual confounding by controlling for temporal trends at the population level. Fourth, we include both oil and gas drilling as well as historical drilling (active wells pre-1996) in our analysis to account for the range of exposures that may occur near a residence. Despite their similar potential for air pollution,[64] exposures beyond unconventional gas drilling have rarely been included in health analyses to date.[15] This set of strengths in our analysis allows our results to considerably expand on existing literature and provide new data on key concerns for local communities.

Although our study has substantial strengths, there are limitations to consider. First, our study is observational in nature and, as such, we cannot interpret our present findings as a causal relationship. Rather, this analysis contributes to the growing body of evidence which suggests that exposures related to oil and gas drilling are associated with adverse health outcomes. Second, our difference-in-differences framework shows some signs of measured population changes that are occurring dissimilarly among our groups such as the proportion of unemployed people and the proportion of Hispanic mothers. This may be due to unmeasured spatial confounding that we do not assess in this study. Although this implementation is not perfect, the difference-in-differences framework still shows that it reduced confounding by demographic changes and socioeconomic status compared with a pre vs post or near vs far study design. Third, we obtained information on hypertensive conditions during pregnancy from birth certificates that are abstracted from the mother's medical record at delivery. This method of outcome ascertainment likely yields an under-reporting of the true incidence of these conditions, particularly for gestational hypertension.[65] We also do not have access to data on pre-eclampsia and it is unclear how those cases may be classified on birth certificates. In this data source, we anticipate that outcome misclassification trends towards under-reporting the diagnoses, which would bias our results towards the null. Fourth, birth-certificate data are unable to provide residential information beyond reported maternal address at delivery. Existing literature estimates that 9–32% of women change residences during pregnancy,[66] but the distances moved may not meaningfully impact environmental exposure assessment.[67] Fifth, we did not account for external sources of air pollution that could confound our analysis such as industrial emissions. Whereas we account for proximity to major roads, more refined co-exposure metrics would improve this analysis. Sixth, drilling activity and infrastructure are inherently more complicated than the coordinates of site locations. Additional components of this infrastructure include pipelines, compressor stations and retention ponds,[25] and other activities that may be occurring at a drilling site include gas flaring, hydraulic fracturing and fluid spills.[15] Although these components could produce local pollution that may increase the risks of hypertension, evaluating each of these exposures is beyond the scope of our present analysis. Seventh, we also note that ancillary exposure sources such as diesel-truck traffic and construction activities likely occur prior to the date that drilling began on the site, which may create some exposure misclassification in our analysis. Finally, we cannot rule out the role that residual confounding may be playing in our results, as with all observational research. We overcome some concerns about residual confounding via the difference-in-differences study design, but it is very possible that other sources of confounding could be infiltrating our results. Our restricted analyses by population characteristics are largely exploratory, thus they should be interpreted with caution. With these limitations in mind, our analysis still provides a novel contribution to the existing literature on how the oil and gas industry may affect population health. Additional work is necessary to confirm our findings in other populations, as this analysis is the first one to our knowledge to specifically examine the association between oil and gas development exposures and hypertensive conditions during pregnancy.

Conclusions

Using a population-based retrospective birth cohort with a large sample size, our study provides the first evidence to date that exposure to oil and gas drilling increases the risks of gestational hypertension and eclampsia. These effects are concentrated among maternal residences at delivery within 1 km of at least one drilling site. Given the substantial burden of hypertension conditions on pregnant women, their families and healthcare systems,[2] associations between oil and gas drilling and elevated risks of gestational hypertension and eclampsia require more research from the scientific community and careful consideration by policymakers.

Supplementary data

Supplementary data are available at *IJE* online.

Ethics approval

This research has been approved by the Texas State Department of Health and Human Services (#15-063) and the Oregon State University Institutional Review Board (#6692).

Funding

This work is partially funded by the National Institute of Environmental Health Sciences, National Institutes of Health [Award Number: F31 ES029801], the National Center for Advancing Translational Sciences, National Institutes of Health [Award Number: TL1 TR002371] and Office of the Director, National Institutes of Health [Award Number: DP5 OD021338]. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data availability

Data are available by request from the Vital Statistics Unit in the Texas Department of State Health Services.

Acknowledgements

We thank Dr Andrew Boslett (University of Rochester) for insightful comments on an earlier version of this manuscript.

Author contributions

M.D.W. conducted the analysis and prepared the first draft of the manuscript. E.L.H. and M.L.K. provided input on the analytical strategy and helped interpret the results. S.C. facilitated access to the health data and helped interpret the results. P.H. supervised the study, provided input on study design and helped interpret the results. All co-authors reviewed and contributed to the writing of the manuscript.

Conflict of interest

None declared.

References

1. Report of the National High Blood Pressure Education Program Working Group on high blood pressure in pregnancy. *Am J Obstet Gynecol* 2000;**183**:s1–22.
2. American College of Obstetrics and Gynecologists. ACOG Practice Bulletin No. 202: gestational hypertension and pre-eclampsia. *Obstet Gynecol* 2019;**133**:1.
3. Stevens W, Shih T, Incerti D *et al.* Short-term costs of preeclampsia to the United States health care system. *Am J Obstet Gynecol* 2017;**217**:237–48.e16.
4. Fingar KR, Mabry-Hernandez I, Ngo-Metzger Q, Wolff T, Steiner CA, Elixhauser A. *Delivery Hospitalizations Involving Preeclampsia and Eclampsia, 2005-2014 Statistical Brief #222.* Rockville, MD: Healthcare Cost and Utilization Project (HCUP) Statistical Briefs, 2006.
5. Wallis AB, Saftlas AF, Hsia J, Atrash HK. Secular trends in the rates of preeclampsia, eclampsia, and gestational hypertension, United States, 1987–2004. *Am J Hypertens* 2008;**21**:521–26.
6. Shen M, Smith GN, Rodger M, White RR, Walker MC, Wen SW. Comparison of risk factors and outcomes of gestational hypertension and pre-eclampsia. *PLoS One* 2017;**12**:e0175914.
7. Brown CM, Garovic VD. Mechanisms and management of hypertension in pregnant women. *Curr Hypertens Rep* 2011;**13**:338–46.
8. Leaños-Miranda A, Méndez-Aguilar F, Ramírez-Valenzuela KL *et al.* Circulating angiogenic factors are related to the severity of gestational hypertension and preeclampsia, and their adverse outcomes. *Medicine (Baltimore)* 2017;**96**:e6005.
9. Wu J, Wilhelm M, Chung J, Ritz B. Comparing exposure assessment methods for traffic-related air pollution in an adverse pregnancy outcome study. *Environ Res* 2011;**111**:685–92.
10. Assibey-Mensah V, Glantz JC, Hopke PK *et al.* Ambient wintertime particulate air pollution and hypertensive disorders of pregnancy in Monroe County, New York. *Environ Res* 2019;**168**:25–31.
11. Rudra CB, Williams MA, Sheppard L, Koenig JQ, Schiff MA. Ambient carbon monoxide and fine particulate matter in relation to preeclampsia and preterm delivery in Western Washington State. *Environ Health Perspect* 2011;**119**:886–92.

12. Savitz DA, Elston B, Bobb JF *et al.* Ambient fine particulate matter, nitrogen dioxide, and hypertensive disorders of pregnancy in New York City. *Epidemiology* 2015;**26**:748–57.
13. National Toxicology Program. *NTP Monograph on the Systematic Review of Traffic-Related Air Pollution and Hypertensive Disorders of Pregnancy*. National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 2019.
14. McKenzie LM, Witter RZ, Newman LS, Adgate JL. Human health risk assessment of air emissions from development of unconventional natural gas resources. *Sci Total Environ* 2012;**424**:79–87.
15. HEI Energy Research Committee. *Human Exposure to Unconventional Oil and Gas Development: A Literature Survey for Research Planning*. Health Effects Institute – Energy, 2019. https://hei-energy.org/publication/human-exposure-unconventional-oil-and-gas-development-literature-survey-research (June 2020, date last accessed).
16. Korfmacher KS, Jones WA, Malone SL, Vinci LF. Public health and high volume hydraulic fracturing. *New Solut J Solut* 2013;**23**:13–31.
17. Adgate JL, Goldstein BD, McKenzie LM. Potential public health hazards, exposures and health effects from unconventional natural gas development. *Environ Sci Technol* 2014;**48**:8307–20.
18. Hill E, Ma L. Shale gas development and drinking water quality. *Am Econ Rev* 2017;**107**:522–25.
19. Bonetti P, Leuz C, Michelon G. Large-sample evidence on the impact of unconventional oil and gas development on surface waters. *Science* 2021;**373**:896–902.
20. Allen DT. Emissions from oil and gas operations in the United States and their air quality implications. *J Air Waste Manag Assoc* 2016;**66**:549–75.
21. Eapi GR, Sabnis MS, Sattler ML. Mobile measurement of methane and hydrogen sulfide at natural gas production site fence lines in the Texas Barnett Shale. *J Air Waste Manag Assoc* 2014;**64**:927–44.
22. Caron-Beaudoin É, Valter N, Chevrier J, Ayotte P, Frohlich K, Verner M-A. Gestational exposure to volatile organic compounds (VOCs) in Northeastern British Columbia, Canada: a pilot study. *Environ Int* 2018;**110**:131–38.
23. Caron-Beaudoin É, Bouchard M, Wendling G *et al.* Urinary and hair concentrations of trace metals in pregnant women from Northeastern British Columbia, Canada: a pilot study. *J Expo Sci Environ Epidemiol* 2019;**29**:613–23.
24. Measham TG, Fleming DA. Assess benefits and costs of shale energy. *Nature* 2014;**510**:473.
25. Czolowski E, Santoro R, Srebotnjak T, Shonkoff S. Toward consistent methodology to quantify populations in proximity to oil and gas development: a national spatial analysis and review. *Environ Health Perspect* 2017;**125**:086004–11.
26. Brown JP, Fitzgerald T, Weber JG. Does resource ownership matter? Oil and gas royalties and the income effect of extraction. *J Assoc Environ Resour Econ* 2019;**6**:1039–64.
27. Weber JG. The effects of a natural gas boom on employment and income in Colorado, Texas, and Wyoming. *Energy Econ* 2012;**34**:1580–88.
28. Feyrer J, Mansur ET, Sacerdote B. Geographic dispersion of economic shocks: evidence from the fracking revolution. *Am Econ Rev* 2017;**107**:1313–34.
29. Maniloff P, Mastromonaco R. The local employment impacts of fracking: a national study. *Resour Energy Econ* 2017;**49**:62–85.
30. Silva TLS, Morales-Torres S, Castro-Silva S, Figueiredo JL, Silva AMT. An overview on exploration and environmental impact of unconventional gas sources and treatment options for produced water. *J Environ Manage* 2017;**200**:511–29.
31. Zwickl K. The demographics of fracking: a spatial analysis for four U.S. states. *Ecol Econ* 2019;**161**:202–15.
32. Rasmussen S, Ogburn E, McCormack M *et al.* Association between unconventional natural gas development in the Marcellus Shale and asthma exacerbations. *JAMA Intern Med* 2016;**176**:1334–43.
33. Willis MD, Jusko TA, Halterman JS, Hill EL. Unconventional natural gas development and pediatric asthma hospitalizations in Pennsylvania. *Environ Res* 2018;**166**:402–08.
34. Willis M, Hystad P, Denham A, Hill E. Natural gas development, flaring practices and paediatric asthma hospitalizations in Texas. *Int J Epidemiol* 2020;**49**:1883–96.
35. Casey JA, Wilcox HC, Hirsch AG, Pollak J, Schwartz BS. Associations of unconventional natural gas development with depression symptoms and disordered sleep in Pennsylvania. *Sci Rep* 2018;**8**:11375.
36. Casey JA, Goin DE, Rudolph KE *et al.* Unconventional natural gas development and adverse birth outcomes in Pennsylvania: the potential mediating role of antenatal anxiety and depression. *Environ Res* 2019;**177**:108598.
37. Boslett A, Hill E, Ma L, Zhang L. Rural light pollution from shale gas development and associated sleep and subjective well-being. *Resour Energy Econ*. 2021;**64**:101220.
38. Casey JA, Savitz DA, Rasmussen SG *et al.* Unconventional natural gas development and birth outcomes in Pennsylvania, USA. *Epidemiology* 2016;**27**:163–72.
39. McKenzie LM, Guo R, Witter RZ, Savitz DA, Newman LS, Adgate JL. Birth outcomes and maternal residential proximity to natural gas development in rural Colorado. *Environ Health Perspect* 2014;**122**:412–17.
40. Hill EL. Shale gas development and infant health: evidence from Pennsylvania. *J Health Econ* 2018;**61**:134–50.
41. Currie J, Greenstone M, Meckel K. Hydraulic fracturing and infant health: new evidence from Pennsylvania. *Sci Adv* 2017;**3**:e1603021.
42. Whitworth K, Marshall A, Symanski E. Maternal residential proximity to unconventional gas development and perinatal outcomes among a diverse urban population in Texas. *PLoS One* 2017;**12**:e0180966.
43. Gonzalez DJX, Sherris AR, Yang W *et al.* Oil and gas production and spontaneous preterm birth in the San Joaquin Valley, CA: a case–control study. *Environ Epidemiol* 2020;**4**:e099.
44. Tran K, Casey J, Cushing L, Morello-Frosch R. Residential proximity to oil and gas development and birth outcomes in California: a retrospective cohort study of 2006–2015 births. *Environ Health Perspect* 2020;**128**:67001.
45. Cushing L, Vavra-Musser K, Chau K, Franklin M, Johnston J. Flaring from unconventional oil and gas development and birth outcomes in the Eagle Ford Shale in South Texas. *Environ Health Perspect* 2020;**128**:77003.
46. Caron-Beaudoin É, Whitworth KW, Bosson-Rieutort D, Wendling G, Liu S, Verner M-A. Density and proximity to hydraulic fracturing wells and birth outcomes in Northeastern British Columbia, Canada. *J Expo Sci Environ Epidemiol* 2021;**31**:53–59.

47. Willis MD, Hill EL, Boslett A, Kile ML, Carozza SE, Hystad P. Associations between residential proximity to oil and gas drilling and term birth weight and small for gestational age infants in Texas: a difference-in-differences analysis. *Environ Health Perspect* 2021;**129**:077002.
48. McKenzie LM, Crooks J, Peel JL *et al.* Relationships between indicators of cardiovascular disease and intensity of oil and natural gas activity in Northeastern Colorado. *Environ Res* 2019;**170**:56–64.
49. Abadie A. Semiparametric difference-in-differences estimators. *Rev Econ Stud* 2005;**72**:1–19.
50. Zhang S, Ding Z, Liu H *et al.* Association between mental stress and gestational hypertension/preeclampsia: a meta-analysis. *Obstet Gynecol Surv* 2013;**68**:825–34.
51. Denham A, Willis M, Zavez A, Hill E. Unconventional natural gas development and hospitalizations: evidence from Pennsylvania, United States, 2003-2014. *Public Health* 2019;**168**:17–25.
52. Gearhart R, Michieka N. Natural resource abundance and healthcare efficiency in Appalachia: a robust conditional approach. *Energy Policy* 2019;**129**:985–996.
53. U.S. Energy Information Administration. *Natural Gas Explained: Where Our Natural Gas Comes From*. 2017. (29 November 2021, date last accessed).
54. Enverus Drillinginfo. https://www.enverus.com.
55. Vandenbroucke JP, von Elm E, Altman DG *et al.*; for the STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med* 2007;**4**:e297.
56. Texas Department of State Health Services. *Vital Statistics Annual Report, Table 44*. 2015. https://www.dshs.texas.gov/chs/vstat/vs08/t44.shtm (17 March 2021, date last accessed).
57. Noris M, Perico N, Remuzzi G. Mechanisms of disease: preeclampsia. *Nat Clin Pract Nephrol* 2005;**1**:98–114.
58. Department of Energy. What Is Environmental Justice? 2016. https://www.energy.gov/lm/services/environmental-justice/what-environmental-justice (29 June 2018, date last accessed).
59. Lee C. Environmental justice: building a unified vision of health and the environment. *Environ Health Perspect* 2002;**110**:141–44.
60. Clark LP, Millet DB, Marshall JD. National patterns in environmental injustice and inequality: outdoor NO2 air pollution in the United States. *PLoS One* 2014;**9**:e94431.
61. Johnston JE, Werder E, Sebastian D. Wastewater disposal wells, fracking, and environmental injustice in Southern Texas. *Am J Public Health* 2016;**106**:550–56.
62. Ghosh G, Grewal J, Männistö T *et al.* Racial/ethnic differences in pregnancy-related hypertensive disease in nulliparous women. *Ethn Dis* 2014;**24**:283–89.
63. Rahm D. Regulating hydraulic fracturing in shale gas plays: the case of Texas. *Energy Policy* 2011;**39**:2974–81.
64. U.S. Department of Energy. *Modern Shale Gas Development in the United States: A Primer*. 2009. https://energy.gov/fe/downloads/modern-shale-gas-development-united-states-primer (19 January 2018, date last accessed).
65. Roberts CL, Bell JC, Ford JB, Hadfield RM, Algert CS, Morris JM. The accuracy of reporting of the hypertensive disorders of pregnancy in population health data. *Hypertens Pregnancy* 2008;**27**:285–97.
66. Bell ML, Belanger K. Review of research on residential mobility during pregnancy: consequences for assessment of prenatal environmental exposures. *J Expo Sci Environ Epidemiol* 2012;**22**:429–38.
67. Lupo P, Symanski E, Chan W *et al.* Differences in exposure assignment between conception and delivery: the impact of maternal mobility. *Paediatr Perinat Epidemiol* 2010;**24**:200–08.

ATTACHMENT C

STUDY 14

Associations between Unconventional Natural Gas Development and Nasal and Sinus, Migraine Headache, and Fatigue Symptoms in Pennsylvania

Aaron W. Tustin,[1] Annemarie G. Hirsch,[2] Sara G. Rasmussen,[1] Joan A. Casey,[3] Karen Bandeen-Roche,[4] and Brian S. Schwartz[1,2,5]

[1]Department of Environmental Health Sciences, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA; [2]Center for Health Research, Geisinger Health System, Danville, Pennsylvania, USA; [3]Robert Wood Johnson Health and Society Scholars Program, University of California, San Francisco, San Francisco and University of California, Berkeley, Berkeley, California, USA; [4]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA; [5]Department of Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland, USA

Background: Unconventional natural gas development (UNGD) produces environmental contaminants and psychosocial stressors. Despite these concerns, few studies have evaluated the health effects of UNGD.

Objectives: We investigated associations between UNGD activity and symptoms in a cross-sectional study in Pennsylvania.

Methods: We mailed a self-administered questionnaire to 23,700 adult patients of the Geisinger Clinic. Using standardized and validated questionnaire items, we identified respondents with chronic rhinosinusitis (CRS), migraine headache, and fatigue symptoms. We created a summary UNGD activity metric that incorporated well phase, location, total depth, daily gas production and inverse distance–squared to patient residences. We used logistic regression, weighted for sampling and response rates, to assess associations between quartiles of UNGD activity and outcomes, both alone and in combination.

Results: The response rate was 33%. Of 7,785 study participants, 1,850 (24%) had current CRS symptoms, 1,765 (23%) had migraine headache, and 1,930 (25%) had higher levels of fatigue. Among individuals who met criteria for two or more outcomes, adjusted odds ratios for the highest quartile of UNGD activity compared with the lowest were [OR (95% CI)] 1.49 (0.78, 2.85) for CRS plus migraine, 1.88 (1.08, 3.25) for CRS plus fatigue, 1.95 (1.18, 3.21) for migraine plus fatigue, and 1.84 (1.08, 3.14) for all three outcomes together. Significant associations were also present in some models of single outcomes.

Conclusions: This study provides evidence that UNGD is associated with nasal and sinus, migraine headache, and fatigue symptoms in a general population representative sample.

Citation: Tustin AW, Hirsch AG, Rasmussen SG, Casey JA, Bandeen-Roche K, Schwartz BS. 2017. Associations between unconventional natural gas development and nasal and sinus, migraine headache, and fatigue symptoms in Pennsylvania. Environ Health Perspect 125:189–197; http://dx.doi.org/10.1289/EHP281

Introduction

Unconventional natural gas development (UNGD), which includes the process of hydraulic fracturing, represents an expanding share of energy production worldwide. Shale gas extraction now comprises 40% of U.S. domestic natural gas production [Energy Information Administration (EIA 2015)]. In the past decade, particularly rapid increases in UNGD have occurred in Pennsylvania, where > 8,800 unconventional wells have been drilled.

There are concerns that UNGD could affect the environment via chemical pollutants such as diesel exhaust, volatile organic compounds, combustion products, fugitive emissions, and fracking chemicals (Werner et al. 2015). UNGD has been linked to contamination of air (Macey et al. 2014; Paulik et al. 2015), soil (Maloney and Yoxtheimer 2012), ground water (Jackson et al. 2013; Drollette et al. 2015), and surface water (Kassotis et al. 2014). UNGD also creates contextual and psychosocial stressors including noise, truck traffic, influxes of nonlocal workers, and perceived negative impacts on quality of life and on the built and social environments (Saberi et al. 2014; Powers et al. 2015; Adgate et al. 2014).

There have been few studies of the health effects of UNGD, despite increasing concerns (Mitka 2012; Kovats et al. 2014). Previous studies have been limited by factors including small sample size and imprecise exposure assessment (Adgate et al. 2014). Because the expansion of UNGD has outpaced scientific understanding of its potential health impacts, studies of self-reported outcomes have been advocated as a rapid means of generating hypotheses that could influence public policy. Furthermore, some illnesses with plausible links to UNGD, such as pain syndromes and fatigue, are defined solely by symptoms. Yet, to date there have been only two epidemiologic studies of symptoms in relation to UNGD, each with < 500 participants (Steinzor et al. 2013; Rabinowitz et al. 2015).

We used data from a large population-based cross-sectional survey of Pennsylvania adults to identify patients with nasal and sinus symptoms, migraine headache, and higher levels of fatigue. We selected these outcomes because of their high prevalence, large economic costs, and possible links to environmental risk factors through chemical toxicity, irritation, odors, or stress (Hastan et al. 2011; Bhattacharyya 2009; Shashy et al. 2004; Tan et al. 2013; Friedman and De ver Dye 2009; Sjöstrand et al. 2010; Bell et al. 1998; Griffith and Zarrouf 2008; Ranjith 2005; Ricci et al. 2007). The purpose of this study was to test the null hypothesis that UNGD is not associated with these three outcomes. To do so, we performed a case–control analysis in which we compared individuals having one or more of these health outcomes with selected participants having no or minimal evidence of these diseases.

Methods

Study Overview

In early 2014, we performed a cross-sectional survey of primary care patients of the Geisinger Clinic. Information was gathered via a questionnaire designed to study general chronic rhinosinusitis (CRS)

Address correspondence to B.S. Schwartz, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe St., Room W7041, Baltimore, MD 21205 USA. Telephone: (410) 955-4158. E-mail: bschwar1@jhu.edu

Supplemental Material is available online (http://dx.doi.org/10.1289/EHP281).

This research was funded by National Institutes of Health grants U19 AI106683 (PI R. Schleimer; A.G.H., B.S.S.), R21 ES023675 (B.S.S.), and training grant ES07141 (S.G.R.); the Robert Wood Johnson Foundation (J.A.C.); the Degenstein Foundation; and the National Science Foundation Integrative Graduate Education and Research Traineeship (S.G.R.).

B.S.S. is a Fellow of the Post Carbon Institute (PCI), serving as an informal advisor on climate, energy, and health issues. He receives no payment for this role. His research is entirely independent of PCI, and is not motivated, reviewed, or funded by PCI.

The other authors declare they have no actual or potential competing financial interests.

Received: 22 December 2015; Revised: 29 July 2016; Accepted: 31 July 2016; Published: 25 August 2016.

epidemiology (for the questionnaire, see Supplemental Material, "Population Study of Nasal and Sinus Symptoms"). The questionnaire did not mention UNGD because that was not its primary purpose. We used residential addresses and information about Pennsylvania unconventional gas wells to create UNGD activity metrics for four time-varying well development phases. We evaluated the associations between UNGD activity and CRS, migraine headache, and fatigue symptoms. The study protocol was approved by the Institutional Review Board (IRB) of the Geisinger Health System with an IRB Authorization Agreement with the Johns Hopkins Bloomberg School of Public Health. Waivers of Health Insurance Portability and Accountability Act of 1996 (HIPAA) authorization and written informed consent were approved by the IRB; implied consent was considered to have been provided if the patient returned the mailed questionnaire.

Study Population

The Geisinger Clinic provides primary care services to > 400,000 patients, predominantly in central and northeastern Pennsylvania. Our source population consisted of 200,769 adult (age ≥ 18 years) Geisinger primary care patients for whom we had electronic health record (EHR) data and information on race/ethnicity. From this source population, we selected 23,700 survey recipients using a stratified sampling design that is described in "Rationale and Description of the Stratified Sampling Method." We mailed the baseline questionnaire in April 2014. A total of 7,847 (33.1%) individuals returned the questionnaire after three mailings. Questionnaires were returned between 13 April and 13 October 2014. After excluding respondents who lived outside Pennsylvania (n = 62), the study sample consisted of 7,785 participants.

Rationale and Description of the Stratified Sampling Method

We oversampled racial/ethnic minorities because a primary interest of the parent grant was to understand racial/ethnic differences in CRS epidemiology. Geisinger's catchment area only has ~8% racial/ethnic minorities. Oversampling was necessary to ensure a sufficient number of racial/ethnic minorities in the parent study.

Similarly, to ensure an adequate number of CRS patients in the parent CRS study, we oversampled individuals with higher likelihood of having CRS. To do so, we used EHR data to identify Geisinger primary care patients with higher, intermediate, and lower likelihood of CRS. These assessments were based on *International Classification of Diseases, Ninth Revision* (ICD-9) codes and Current Procedural Terminology (CPT) codes from the medical record. Patients with a "higher" likelihood of CRS (n = 13,494) had at least two ICD-9 codes for CRS (ICD-9 codes 473.x or 471.x) associated with an outpatient, inpatient, or emergency department encounter, or they had at least one CPT code for sinus computerized tomography, sinus endoscopy, or sinus surgery. Patients with "intermediate" likelihood of CRS (n = 49,918) had at least one ICD-9 code for asthma (493.x) or allergic rhinitis (477.x) or a single ICD-9 code for CRS associated with an outpatient, inpatient, or emergency department encounter. The 137,357 patients who did not meet criteria for the higher and intermediate likelihood groups were designated as having a "lower" likelihood of CRS.

We divided our source population into six strata based on race/ethnicity and likelihood of CRS. We mailed the baseline CRS survey to a larger percentage of individuals in the strata of interest (see Table S1).

Covariates

We obtained the following covariates from the EHR: sex, current age (years), race/ethnicity (white non-Hispanic, other), smoking status (never, current, former), body mass index (BMI; kg/m^2), residential address, and history of receiving Medical Assistance, a means-tested health insurance program that we used as a surrogate for family socioeconomic status (Casey et al. 2013). We used information in the EHR to derive each individual's residential place type (township, borough, or census tract in cities) and Charlson comorbidity index. We computed the Charlson index, which incorporates the number and severity of comorbid illnesses, in a manner consistent with previously published criteria (Charlson et al. 1987). We dichotomized race/ethnicity because only 10% of participants were nonwhite, which is reflective of the general population in these communities (Casey et al. 2016). Our questionnaire ascertained additional information on educational status, marital status, household income, hay fever, nasal polyps, age at onset of nasal/sinus symptoms (in 5-year categories), history of sinus surgery, and current use of sinusitis medications (antibiotics and oral, inhaled, and nasal corticosteroids). We used U.S. Census data (Liu et al. 2012) to derive community socioeconomic deprivation (CSD) in townships, boroughs, and cities using a modified version of the Townsend index (Townsend 1987) as previously reported (Liu et al. 2012).

Outcome Ascertainment

The cardinal symptoms of CRS are nasal congestion/obstruction, nasal discharge (anterior or posterior nasal drip), smell loss, and facial pain or pressure. Our questionnaire ascertained the frequency ("never," "once in a while," "some of the time," "most of the time," or "all the time"), in the past 3 months, of the aforementioned symptoms (questions 10–15 of the questionnaire, which is included in the Supplemental Material, "Population Study of Nasal and Sinus Symptoms"). Following European Position Paper on Rhinosinusitis and Nasal Polyps (EPOS) diagnostic criteria for CRS in epidemiologic studies (Fokkens et al. 2012), we determined participants to have current CRS if they experienced two or more cardinal symptoms [one of which must be nasal congestion/obstruction (question 10) or discharge (question 11 and/or 12)] at least "most of the time" in the past 3 months.

We ascertained migraine headache via questions from the ID Migraine™ questionnaire (Lipton et al. 2003) covering the past 12 months. Those with headaches at least "some of the time" (question 80) were asked the frequency ("never," "rarely," "less than half the time," "half the time or more") of headache-associated disability, nausea, and photophobia (questions 81–83). Using a validated scoring method (Lipton et al. 2003), we dichotomized the three responses. Responses of "never" or "rarely" were scored as no and responses of "less than half the time" or "half the time or more" were scored as yes. Participants who answered yes to at least two of three questions were considered to have migraines.

We ascertained fatigue with eight questions from the Patient-Reported Outcomes Measurement Information System (PROMIS®) fatigue short form 8a (http://www.assessmentcenter.net). These items assessed the frequency ("not at all," "a little bit," "somewhat," "quite a bit," "very much") of fatigue and fatigue-related disability in the past week (questions 84–91). We used the instrument's standardized scoring instructions to code responses from 1 ("not at all") to 5 ("very much") and summed the eight values to produce a score ranging from 8 to 40. We excluded individuals who answered fewer than four questions (n = 76). Individuals who answered between 4 and 7 questions were assigned a prorated score using the following formula: score = (raw sum × 8)/(number of items answered). Fractional scores were rounded up to the nearest integer. Our "higher levels of fatigue" outcome consisted of individuals in the highest quartile (score ≥ 28).

Some respondents met criteria for more than one outcome. In the analysis, we evaluated associations of UNGD with single outcomes (i.e., CRS only, migraine only, or fatigue only) and with multiple outcomes (i.e., participants with CRS and migraine, CRS and fatigue, migraine and fatigue, or all three outcomes).

Reference Group

We performed an unmatched case–control analysis in which we compared individuals having one or more of the three primary outcomes ("cases") with a subset of participants having no or minimal evidence of these outcomes (hereafter referred to as "controls" or the "reference group"). The reference group comprised study participants who *a*) did not meet diagnostic criteria for past or current CRS, *b*) reported no migraine headache symptoms, and *c*) reported lower levels of fatigue (i.e., first quartile of fatigue score). Individuals with past CRS, intermediate likelihood of migraine, and/or moderate levels of fatigue were excluded from the reference group. These exclusion criteria were intended to produce a reference group free of individuals with a moderate likelihood of having the outcome (in the case of migraine and fatigue) or whose disease had been aggressively managed and treated (in the case of past CRS).

We created the reference group as follows. First, we excluded all study participants with one or more of the outcomes of interest. Next, individuals who met criteria for lifetime CRS [i.e., responses of "yes" to at least two cardinal symptoms on questions 1–6, one of which had to be nasal blockage (question 1) or discharge (question 2 and/or 3)] but not current CRS were deemed to have "past CRS" and were excluded from the reference group. We then excluded participants from the reference group if they endorsed any of the three ID Migraine™ criteria. In other words, members of the reference group either skipped the ID Migraine™ questions (e.g., because they reported a headache frequency of "never" or "once in a while" on question 80 and were instructed to skip the following three questions) or responded to questions 81–83 with no migraine symptom occurring more frequently than "never" or "rarely." Finally, we excluded individuals from the reference group if their fatigue score was higher than the 25th percentile (i.e., those with fatigue score > 13) or if they did not answer at least four of eight PROMIS® fatigue items (questions 84–91). No other inclusion or exclusion criteria were applied to the reference group.

UNGD Activity Assessment

We used published descriptions, and our own data, to estimate the duration of each UNGD phase (Gaines and Ziegler 2013; New York State Department of Environmental Conservation 2015; Casey et al. 2016). Pad preparation, which involves clearing of the well site, lasts ~30 days. Drilling the well then takes 1–30 days, proportionate to the total (vertical plus horizontal) depth. After drilling, hydraulic fracturing (fracking) occurs during a stimulation phase that lasts an average of 7 days. Finally, the well produces natural gas during a production phase that lasts months to years.

To capture these complexities of well development, we compiled data on UNGD in Pennsylvania from 1 January 2005 through 31 December 2014 from the Pennsylvania Department of Environmental Protection, the Pennsylvania Department of Conservation and Natural Resources, and SkyTruth (http://skytruth.org). We obtained the following information for each well: geographic coordinates; start dates of drilling, stimulation, and production; total depth; and volume of natural gas produced during 6- or 12-month reporting windows.

Using methods described previously (Casey et al. 2016), we created UNGD activity metrics for each phase of well development. Briefly, these metrics incorporated all unconventional gas wells in Pennsylvania and were defined as follows:

$$\text{Metric for participant } i = \frac{1}{T}\sum_{t=-1}^{-T}\sum_{j=1}^{n} w_j(t)\big/d_{ij}^2,$$

where T is an averaging period in days (in our primary analysis, $T = 90$ because CRS diagnostic criteria require 3 months of symptoms); t is a temporal summation index whose negative sign represents past dates (e.g., summing from $t = -1$ to -90 indicates that the metric was averaged over 90 consecutive days immediately before the survey); n is the number of wells; $w_j(t)$ is the weight assigned to the jth well on day t; and d_{ij}^2 is the squared distance between well j and the residential address of participant i. We set $w_j(t) = 0$ for wells that were inactive in the given phase on day t. Active wells were assigned weights during the duration of the relevant phase as follows: for pad preparation and drilling metrics, $w_j(t)$ was 1; for the stimulation metric, $w_j(t)$ was the total depth in feet (a surrogate for hydraulic fracturing chemical volumes and the number of truck trips required to transport stimulation materials); and for the production metric, $w_j(t)$ was the average daily volume in Mcf (1 Mcf = 1,000 cubic feet) of natural gas produced during the corresponding reporting period.

Because the four UNGD phase metrics were highly correlated when averaged over 90 days (Spearman coefficient > 0.90 for each pairwise comparison), we z-transformed the metrics and summed the resulting z-scores. For analysis, we divided this continuous composite UNGD activity metric into quartiles for ease of interpretation and because of its skewed distribution.

Statistical Analysis

We used descriptive statistics to compare characteristics of participants with and without each outcome. To evaluate selection bias with respect to UNGD, we compared distributions of the UNGD activity metric in study participants and questionnaire nonresponders. To assess the potential for nonconservative errors due to selection bias with respect to health status, we analyzed distributions of the Charlson comorbidity index in study participants and survey nonresponders, stratified by UNGD quartile. Categorical and continuous variables were compared using χ^2 tests and t-tests, respectively. For hypothesis testing, p-values < 0.05 were considered statistically significant.

We used weighted logistic regression to evaluate associations between UNGD activity and symptoms while adjusting for confounding variables. All models compared individuals with the outcome(s) of interest ("cases") to the reference group described above ("controls"). The use of sampling weights allowed us to account for the differential patient selection and participation rates in our stratified design while targeting unbiased measures of association and obtaining robust standard errors. We assigned each participant a sampling weight equal to the inverse probability of inclusion in the study (see Table S1). Because the weight in one stratum (150.8) was substantially larger than the other weights, we truncated this weight by reducing it to the value of the second-highest weight (32.3).

We adjusted all models for these potential confounders that we identified *a priori*: sex, race/ethnicity (non-Hispanic white vs. other), age [linear and quadratic terms; to avoid collinearity, we centered the age variable by subtracting its mean (i.e., $A_c = A_i - A_{\text{mean}}$)], receipt of Medical Assistance (never vs. ever), and smoking status (never vs. former and current). We tested for additional confounding by adding linear and quadratic terms for BMI and CSD. We retained these covariates in the models if they changed associations between UNGD and the outcome by ≥ 10%. Analyses were performed in R (version 3.0.2, R Project for Statistical Computing) and Stata 13.1 (StataCorp) using the svy command.

We reasoned that UNGD might be associated with current CRS only for onset of symptoms after 2006, when UNGD commenced in Pennsylvania. To test the associated hypothesis, we stratified the CRS group by date of symptom onset (before/after 1 January 2006) and reran models within each stratum. Although associations of UNGD activity with our other outcomes could also differ by onset date, our questionnaire did not ascertain the onset date of migraine and fatigue symptoms.

We performed several sensitivity analyses. To explore the impact of sampling weight choices, we reran models with full (i.e., not truncated) weights and again with no weights. To determine whether associations differed by the length of the UNGD assessment period, we compared associations using 7-day, 90-day, and 365-day averaged UNGD metrics that corresponded to the questionnaire's recall windows for the three primary outcomes. To explore spatial differences among groups of participants, we mapped the residential locations of individuals with and without our primary outcomes stratified by UNGD quartile and case/control status. To assess whether UNGD was associated with symptoms in individuals with past disease or moderate symptoms, we created additional CRS and fatigue models in which we reclassified some previously excluded individuals as "cases" (for details see Supplemental Material, "Models of Past Disease and Moderate Symptoms"). To assess whether unmeasured confounding, including spatial confounding, could be responsible for the observed associations, we created "negative control outcome" models (Lipsitch et al. 2010). These adjusted logistic regression models evaluated associations between UNGD and self-reported outcomes (bad breath, ear pain, and cold/flu symptoms) that we thought were unlikely to be related to UNGD. We expected to find no significant associations between UNGD and these outcomes; the presence of such associations could indicate bias resulting from unmeasured confounding. In these models, we defined cases as all study participants who reported the symptom at least "most of the time" in the past 3 months (questions 36, 43, and 48 for bad breath, ear pain, and cold/flu symptoms, respectively). The reference group for each model consisted of all individuals who reported the symptom "never" in the past 3 months.

Results

Characteristics of the Study Population

Questionnaire respondents were 7,785 individuals from 39 counties in central and northeastern Pennsylvania, in regions with and without UNGD (Figure 1). Compared with questionnaire recipients who did not respond, our study population was more likely to be female, white, and older (results not shown). The continuous UNGD activity metric did not differ significantly ($p = 0.26$) between study participants and questionnaire nonresponders (Table 1). Study participants were less likely than nonresponders to be in the highest UNGD quartile. Although the Charlson comorbidity index was higher in responders (mean = 3.43) than in nonresponders (mean = 2.52, $p < 0.001$), the mean Charlson values were similar across all UNGD quartiles (Table 1).

We identified 738 participants with current CRS and no other primary outcome, 580 with migraine headache only, and 666 with higher levels of fatigue only (Table 2). These conditions were co-occurring in other individuals. There were 268 individuals with CRS and migraine, 347 with CRS and higher levels of fatigue, 420 with migraine and higher levels of fatigue, and 497 with all three outcomes. There were 1,380 participants with no current or past CRS, no migraine headache symptoms, and lower levels of fatigue; these individuals comprised the reference group. Compared with the reference group, individuals with each single outcome were more likely to be younger and current smokers (Table 2). Those with migraine and fatigue were more likely to be female, and those reporting CRS and fatigue were more likely to be white non-Hispanic.

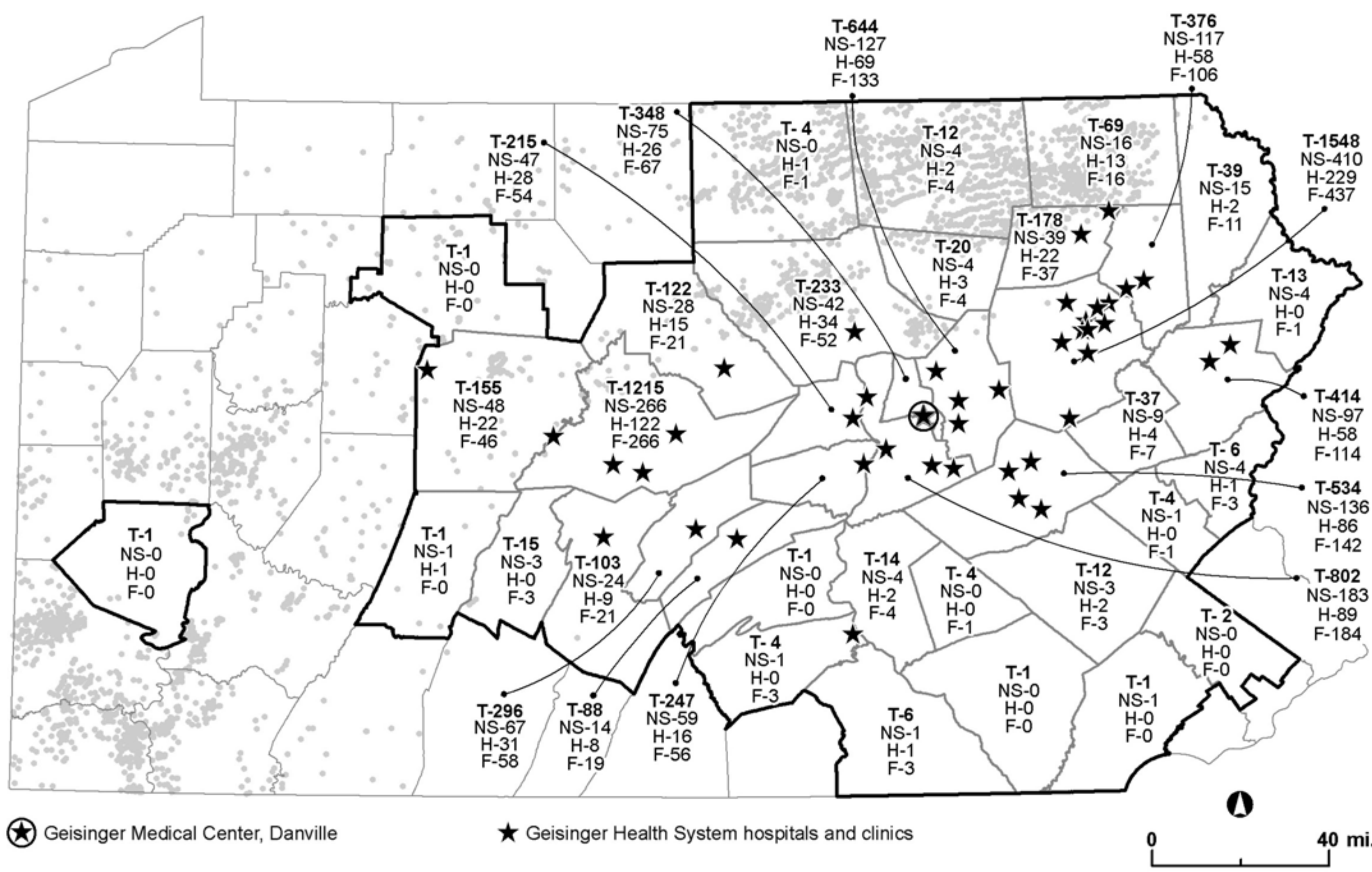


Figure 1. Map of study area. Thick black outlines designate Pennsylvania counties with at least one participant [from U.S. Census Bureau TIGER/line files (U.S. Census Bureau 2010)]. Numbers within the borders of each county indicate the total number of participants (T) and the number with chronic rhinosinusitis symptoms (NS), migraine headache (H), and higher levels of fatigue (F) (data from the Geisinger Clinic). Gray circles show locations of drilled unconventional natural gas wells as of December 2014 (Pennsylvania Department of Environmental Protection 2016). Black stars represent Geisinger hospitals and clinics. Map was made with ArcGIS Desktop (release 10, Esri, Redlands, CA).

Associations of UNGD with Symptoms

The highest quartile of UNGD activity, compared with the lowest, was associated with significantly increased odds of the following combinations of two or more outcomes: CRS and higher levels of fatigue [odds ratio (OR) = 1.88; 95% confidence interval (CI): 1.08, 3.25], migraine headache and higher levels of fatigue (OR = 1.95; 95% CI: 1.18, 3.21), and all three outcomes (OR = 1.84; 95% CI: 1.08, 3.14) (Table 3). The second and third quartiles of UNGD were not significantly associated with any of the outcomes. In individuals with only one outcome, the odds ratios for the fourth quartile of UNGD were 1.11 (95% CI: 0.75, 1.65) for current CRS, 1.43 (95% CI: 0.94, 2.18) for migraine headache, and 1.47 (95% CI: 0.996, 2.18) for higher levels of fatigue (Table 3). In general, participants in the fourth quartile of UNGD lived farther north than those in other UNGD quartiles (Figure 2).

When we stratified CRS patients by onset date, the second (OR = 3.27; 95% CI: 1.21, 8.82) and fourth (OR = 3.26; 95% CI: 1.14, 9.36) quartiles of UNGD were associated with significantly increased odds of CRS in those whose symptoms began after 2006 (see Table S2). There were no associations in participants with earlier symptom onset.

Sensitivity Analyses

In participants with multiple outcomes, most inferences were unchanged whether we used the full sampling weights, truncated

Table 1. Comparison of selected characteristics in survey responders and nonresponders.

Characteristic	Responders ($n = 7{,}785$)	Nonresponders ($n = 15{,}525$)	p-Value
Continuous composite UNGD activity metric, mean ± sd	–0.02 ± 1.80	0.01 ± 2.78	0.26[a]
UNGD activity, n (%)			
Quartile 1	2,052 (26.4)	3,775 (24.3)	< 0.001[b]
Quartile 2	1,828 (23.5)	3,996 (25.7)	
Quartile 3	2,017 (25.9)	3,814 (24.6)	
Quartile 4	1,888 (24.3)	3,940 (25.4)	
Charlson index, mean ± sd	3.43 ± 2.76	2.52 ± 2.65	< 0.001[a]
Charlson index stratified by quartiles of UNGD activity, mean ± sd			
Quartile 1	3.27 ± 2.61	2.46 ± 2.46	
Quartile 2	3.37 ± 2.71	2.48 ± 2.57	NA
Quartile 3	3.61 ± 2.83	2.68 ± 2.85	
Quartile 4	3.47 ± 2.86	2.48 ± 2.70	
	$p < 0.001$[c]	$p < 0.001$[c]	

Notes: NA, not applicable; sd, standard deviation; UNGD, unconventional natural gas development.
Patients who lived outside Pennsylvania were excluded ($n = 390$). UNGD activity was averaged over 90 days before the survey.
[a] p-Value computed using Student's t-test.
[b] p-Value computed using Chi-squared test.
[c] Within responders and nonresponders separately, p-values were computed with one-way analysis of variance (ANOVA) to compare mean Charlson index across quartiles of UNGD.

Table 2. Characteristics of study population by self-reported outcome(s).

		Individuals with no primary outcome		Individuals with one or more primary outcomes						
Characteristic	Overall study population	Reference group[a]	Individuals who were neither cases nor controls[b]	Current CRS only	Migraine headache only	Higher levels of fatigue only	Current CRS and migraine	Current CRS and higher levels of fatigue	Migraine and higher levels of fatigue	Current CRS, migraine headache, and higher levels of fatigue
Total number, n	7,785	1,380	2,889	738	580	666	268	347	420	497
Sex, n (%)										
Male	2,909 (37.4)	656 (47.5)	1,242 (43.0)	335 (45.4)	113 (19.5)	233 (35.0)	50 (18.7)	126 (36.3)	63 (15.0)	91 (18.3)
Female	4,876 (62.6)	724 (52.5)	1,647 (57.0)	403 (54.6)	467 (80.5)	433 (65.0)	218 (81.3)	221 (63.7)	357 (85.0)	406 (81.7)
Race/ethnicity, n (%)										
White non-Hispanic	7,043 (90.5)	1,183 (85.7)	2,653 (91.8)	707 (95.8)	508 (87.6)	598 (89.8)	257 (95.9)	333 (96.0)	357 (85.0)	447 (89.9)
Other	742 (9.5)	197 (14.3)	236 (8.2)	31 (4.2)	72 (12.4)	68 (10.2)	11 (4.1)	14 (4.0)	63 (15.0)	50 (10.1)
Age in years, mean ± sd	55.3 ± 16.1	58.8 ± 17.0	57.6 ± 15.9	57.1 ± 14.9	46.1 ± 14.3	57.3 ± 15.1	48.5 ± 13.2	56.1 ± 14.7	46.5 ± 13.6	47.8 ± 13.1
Smoking status, n (%)										
Never	4,268 (54.8)	805 (58.3)	1,615 (55.9)	404 (54.7)	340 (58.6)	334 (50.2)	141 (52.6)	178 (51.3)	220 (52.4)	231 (46.5)
Current	1,130 (14.5)	134 (9.7)	353 (12.2)	100 (13.6)	96 (16.6)	113 (17.0)	57 (21.3)	61 (17.6)	86 (20.5)	130 (26.2)
Former	2,387 (30.7)	441 (32.0)	921 (31.9)	234 (31.7)	144 (24.8)	219 (32.9)	70 (26.1)	108 (31.1)	114 (27.1)	136 (27.4)
History of receiving medical assistance, n (%)										
Never	6,876 (88.3)	1,286 (93.2)	2,690 (93.1)	694 (94.0)	467 (80.5)	588 (88.3)	216 (80.6)	293 (84.4)	302 (71.9)	340 (68.4)
Ever	909 (11.7)	94 (6.8)	199 (6.9)	44 (6.0)	113 (19.5)	78 (11.7)	52 (19.4)	54 (15.6)	118 (28.1)	157 (31.6)
Body mass index (kg/m^2), mean ± sd	30.2 ± 7.0	29.0 ± 6.3	29.9 ± 6.5	30.4 ± 7.0	29.7 ± 7.3	31.7 ± 7.9	29.8 ± 7.3	31.3 ± 7.4	31.7 ± 7.7	31.2 ± 8.1
Place type, n (%)										
Township	4,949 (63.6)	907 (65.7)	1,900 (65.8)	476 (64.5)	332 (57.2)	417 (62.6)	170 (63.4)	213 (61.4)	242 (57.6)	292 (58.8)
Borough	2,135 (27.4)	371 (26.9)	762 (26.4)	188 (25.5)	183 (31.6)	192 (28.8)	72 (26.9)	101 (29.1)	122 (29.0)	144 (29.0)
Census tract in city	701 (9.0)	102 (7.4)	227 (7.9)	74 (10.0)	65 (11.2)	57 (8.6)	26 (9.7)	33 (9.5)	56 (13.3)	61 (12.3)
Community socioeconomic deprivation, mean ± sd	0.0 ± 3.6	–0.3 ± 3.6	–0.1 ± 3.6	–0.1 ± 3.5	0.3 ± 3.7	0.1 ± 3.6	0.2 ± 3.5	0.1 ± 3.7	0.6 ± 3.7	0.6 ± 3.8
UNGD activity metric, n (%)[c]										
Quartile 1 [–0.61 to –0.47]	1,946 (25.0)	358 (25.9)	745 (25.8)	181 (24.5)	140 (24.1)	155 (23.3)	63 (23.5)	91 (26.2)	101 (24.0)	112 (22.5)
Quartile 2 [–0.47 to –0.39]	1,946 (25.0)	345 (25.0)	731 (25.3)	187 (25.3)	145 (25.0)	174 (26.1)	65 (24.3)	83 (23.9)	92 (21.9)	124 (24.9)
Quartile 3 [–0.39 to –0.16]	1,946 (25.0)	373 (27.0)	733 (25.4)	188 (25.5)	131 (22.6)	172 (25.8)	70 (26.1)	73 (21.0)	98 (23.3)	108 (21.7)
Quartile 4 [> –0.16]	1,947 (25.0)	304 (22.0)	680 (23.5)	182 (24.7)	164 (28.3)	165 (24.8)	70 (26.1)	100 (28.8)	129 (30.7)	153 (30.8)

Notes: CRS, chronic rhinosinusitis; sd, standard deviation; UNGD, unconventional natural gas development.
Percentages may not total 100 because of rounding.
[a] Individuals in the reference group reported no past or current CRS; no headache-related nausea, photophobia, or disability; and lower levels (≤ 25th percentile) of fatigue.
[b] These individuals did not meet criteria for any primary outcome and were excluded from the reference group because of past CRS, intermediate probability of migraine headache, moderate levels of fatigue, or a combination of any of these symptoms.
[c] UNGD activity was averaged over the 90 days before the survey.

weights, or no weights (compare Table 3 with Table S3). Odds ratios for the fourth quartile of UNGD were consistently higher, and had wider confidence intervals, in fully weighted models than in models with truncated weights. For example, the odds ratio for the association of the fourth quartile of UNGD with the coexistence of migraine and fatigue was 2.89 (95% CI: 1.45, 5.76) in the fully weighted model. In individuals with single outcomes, the fourth quartile of UNGD was significantly associated with migraine headache (OR = 1.80; 95% CI: 1.02, 3.17) and fatigue (OR = 1.89; 95% CI: 1.10, 3.26) in the models with full weights; significant associations were also present in unweighted models (see Table S3).

UNGD activity, when averaged over 7 or 365 days, was highly correlated with the 90-day time-averaged UNGD metric used in the primary analyses (Spearman coefficient = 0.98 for both comparisons). Most inferences and associations were similar when using a 7-day or 365-day averaging period (see Table S4). The second quartile of UNGD was associated with past CRS, but there were no associations of UNGD with moderate levels of fatigue (see Table S5). UNGD was not associated with the negative control outcomes of ear pain, bad breath, or cold/flu symptoms (Table 4).

Because only the highest level of UNGD was associated with our primary outcomes, we compared demographic and socioeconomic characteristics of individuals in the fourth quartile of UNGD with those of participants in other UNGD quartiles (see Table S6). Participants in the fourth quartile of UNGD differed on some covariates, several of which were included in the final models. We did not include place type in the final adjusted models because it could be a surrogate for mediators (e.g., individual- or place-level socioeconomic status) of associations between UNGD and symptoms. In a sensitivity analysis that explored the effect of place type, some associations were slightly attenuated when place type was added to the models, but inferences were similar (see Table S7).

Discussion

In our survey of primary care patients in central and northeast Pennsylvania, residential UNGD activity was associated with nasal and sinus symptoms, migraine headache, and higher levels of fatigue, either alone or in combination. Our findings are suggestive of a threshold in the relationship between UNGD and symptoms because associations were present only among participants in the fourth quartile of UNGD activity. We found stronger associations in individuals with two or more co-occurring outcomes. In addition, UNGD was associated with CRS in individuals whose nasal and sinus symptoms began after the start of UNGD in Pennsylvania, although these estimates had lower precision owing to the small number of subjects with recent CRS onset.

In surveys such as ours, in which selection is based on the outcome, regression models must include sampling weights (or employ another strategy to acknowledge the selection mechanism) to avoid bias. However, extreme sampling weights can significantly increase the model's variance (Potter 1988). To balance bias reduction against variance inflation, several techniques have been developed to truncate large sampling weights. We employed one such technique in our primary analyses. We found associations between UNGD and symptoms in the primary models as well as in fully weighted and unweighted models.

There is limited prior evidence linking environmental factors to CRS, migraine headache, and fatigue. Exposure to allergens, toxicants, and secondhand smoke may trigger nasal and sinus symptoms (Fokkens et al. 2012). However, a recent review found insufficient epidemiologic evidence from which to draw conclusions about occupational or environmental risk factors for CRS (Sundaresan et al. 2015). Although migraines have a strong hormonal and genetic component, they can also be triggered by noise, odors, and stress (Friedman and De ver Dye 2009; Sjöstrand et al. 2010; Sauro and Becker 2009). Similarly, fatigue has multiple risk factors including sleep deprivation, psychosocial stressors, medical disorders, psychiatric factors, occupation, and exposure to low levels of environmental chemicals (Bell et al. 1998; Ranjith 2005; Ricci et al. 2007; Griffith and Zarrouf 2008). Our UNGD activity metrics were designed to capture all potential environmental pathways that could affect these symptoms.

We did not measure participants' exposure to ambient air pollution. We also did not account for conventional oil and gas wells. During our study period, the production of conventional gas wells in Pennsylvania was very low compared with that of unconventional wells. Furthermore, Pennsylvania's conventional wells tend to be in the northwest and west, where Geisinger has no clinics/hospitals. The lack of significant geographic overlap with our study population makes confounding of UNGD associations by conventional oil and gas wells unlikely.

Participants in the fourth quartile of UNGD activity lived farther north than those in other quartiles (Figure 2). This spatial separation is due to the location of the Marcellus shale, which constrains UNGD to the northern portion of the Geisinger catchment area. Given the correlation between geography and UNGD, we cannot rule out the possibility that spatial confounding was responsible for the observed associations. However, we note that our models were adjusted for several covariates (such as race/ethnicity and socioeconomic status) that could be associated with both location and outcome. In addition, the null results in our negative control outcome models did not suggest spatial confounding.

CRS, migraine headache, and fatigue are highly prevalent and produce significant societal costs. CRS affects 2–16% of U.S. adults and results in emergency department visits, antibiotic prescriptions, sinus surgeries, and direct healthcare costs (Hastan et al. 2011; Bhattacharyya 2009; Shashy et al. 2004; Tan et al. 2013). Migraines have a prevalence of 11–14% and cause substantial temporary disability, emergency department visits, outpatient clinic visits, and analgesic use (Lipton et al. 2007; Burch et al. 2015). Fatigue

Table 3. Associations of UNGD with symptoms in individuals with one or more primary outcomes, compared with a reference group.

	Adjusted odds ratios (95% confidence intervals)						
UNGD quartile	Current CRS only (n = 736)[a]	Migraine headache only (n = 580)	Higher levels of fatigue only (n = 666)	Current CRS and migraine (n = 266)[a]	Current CRS and higher levels of fatigue (n = 347)[a]	Migraine and higher levels of fatigue (n = 420)	All three outcomes (n = 496)[a]
1	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
2	1.17 (0.80, 1.72)	1.14 (0.74, 1.75)	1.48 (1.01, 2.17)	0.82 (0.43, 1.57)	1.06 (0.62, 1.80)	1.06 (0.63, 1.78)	1.05 (0.63, 1.78)
3	0.76 (0.52, 1.12)	0.89 (0.58, 1.36)	1.22 (0.84, 1.77)	0.74 (0.38, 1.47)	0.94 (0.53, 1.66)	0.80 (0.49, 1.31)	0.73 (0.42, 1.27)
4	1.11 (0.75, 1.65)	1.43 (0.94, 2.18)	1.47 (0.996, 2.18)	1.49 (0.78, 2.85)	1.88 (1.08, 3.25)	1.95 (1.18, 3.21)	1.84 (1.08, 3.14)

Notes: CRS, chronic rhinosinusitis; UNGD, unconventional natural gas development.
For all models, the reference group consisted of individuals with no current or past CRS, no migraine headache symptoms, and the lowest quartile of fatigue score. All models included sampling weights, with the highest weight truncated to the value of the second-highest weight. Models included the following covariates: sex, race/ethnicity (white non-Hispanic vs. other), centered age (linear and quadratic terms), Medical Assistance (never vs. ever), and smoking status (never vs. current and former). UNGD activity was averaged over the 90 days before the survey.
[a]These models included centered body mass index as an additional covariate. Because individuals with unknown body mass index were excluded, these case counts are slightly lower than those reported in the text.

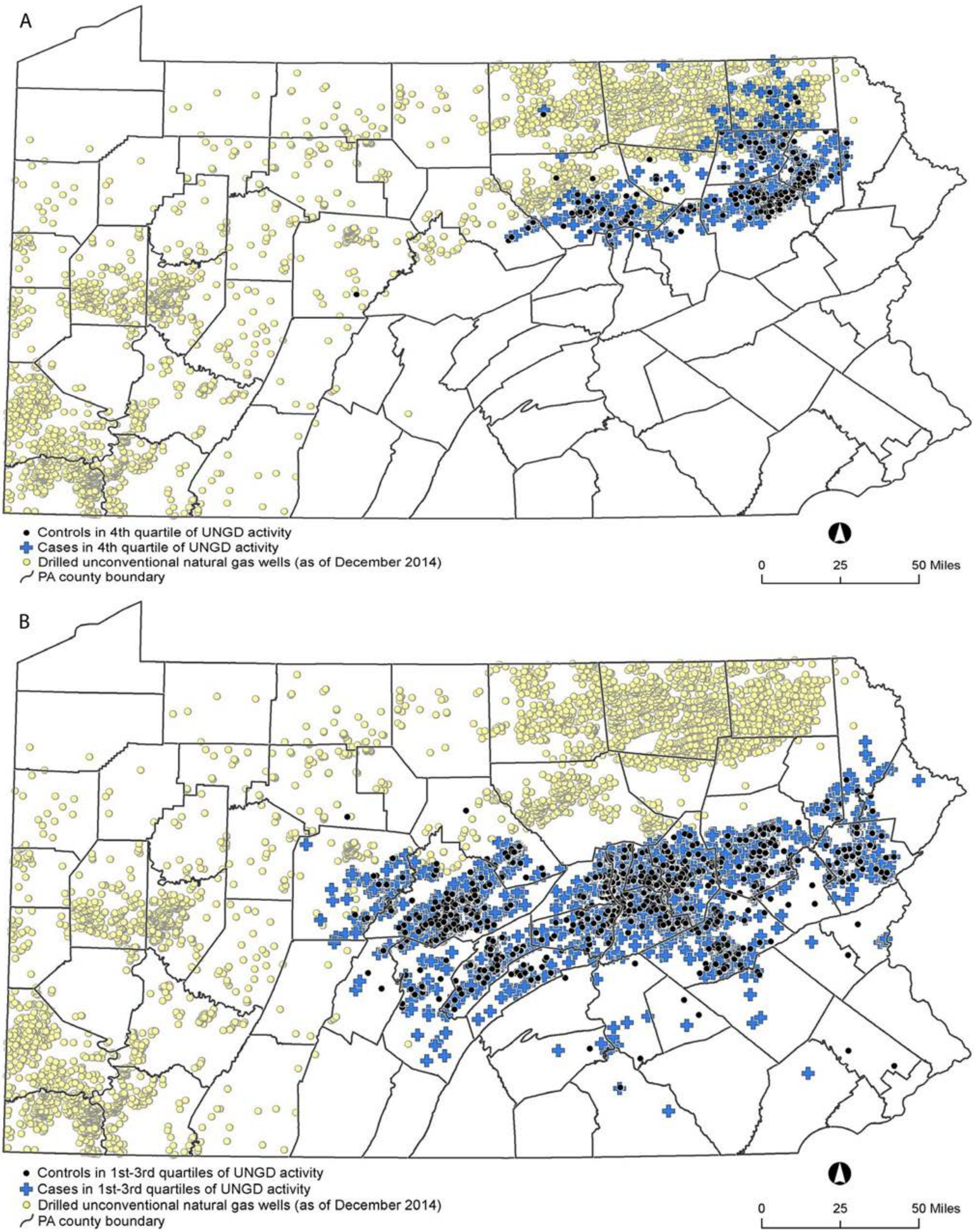


Figure 2. Locations of study participants in the fourth quartile of unconventional natural gas develpment (UNGD) activity (*A*) and all other UNGD quartiles (*B*). Blue crosses: participants with at least one primary outcome [current chronic rhinosinusitis (CRS), migraine headache, and/or higher levels of fatigue]. Black circles: reference group participants with no current or past CRS, no migraine headache symptoms, and lower levels of fatigue. Yellow circles: locations of all drilled unconventional natural gas wells in Pennsylvania as of 31 December 2014. Patient residential locations were from the Geisinger Clinic; county boundaries from the U.S. Census Bureau TIGER/line files (U.S. Census Bureau 2010); and UNGD well locations from the Pennsylvania Department of Environmental Protection (Pennsylvania Department of Environmental Protection 2016). Maps were made with ArcGIS Desktop (release 10, Esri, Redlands, CA).

Table 4. Associations of UNGD with negative control outcomes.

UNGD quartile	Adjusted odds ratios (95% confidence intervals)		
	Ear pain yes (n = 422) vs. no (n = 3,917)	Bad breath yes (n = 846) vs. no (n = 2,628)	Cold/flu symptoms yes (n = 307) vs. no (n = 2,442)
1	1.00 (reference)	1.00 (reference)	1.00 (reference)
2	0.92 (0.58, 1.44)	0.87 (0.61, 1.22)	1.04 (0.58, 1.84)
3	0.53 (0.32, 0.87)	1.12 (0.80, 1.57)	1.15 (0.66, 2.00)
4	1.16 (0.74, 1.83)	0.95 (0.67, 1.35)	1.14 (0.64, 2.01)

UNGD, unconventional natural gas development.
Individuals having the symptom at least "most of the time" in the past 3 months were compared with those having the symptom "never" in the past 3 months. All models included sampling and response weights, and the highest weight was truncated to the value of the second-highest weight. Models included the following covariates: sex, race/ethnicity (white non-Hispanic vs. other), centered age (linear and quadratic terms), Medical Assistance (never vs. ever), and smoking status (never vs. current and former). UNGD activity was averaged over the 90 days before the survey.

prevalence, defined in various ways across studies, is estimated at 7–45%, and fatigue costs U.S. employers > 100 billion USD per year in lost productive work time (Ricci et al. 2007). From a public health and economic perspective, it is vital to understand modifiable risk factors for these illnesses.

Recent reviews have noted the lack of high-quality evidence regarding the health effects of UNGD (Adgate et al. 2014; Werner et al. 2015). Our study of 7,785 Pennsylvania residents is the largest survey of symptoms with respect to UNGD and has several strengths when compared with prior studies. We selected a population-based adult sample with no exclusion criteria. Reporting bias was minimized by the fact that UNGD was not identified as a study aim, and response rates did not differ by proximity to UNGD. Our time-varying UNGD activity metric incorporated well phase and intensity measures such as total depth and gas production. We used standardized and validated instruments to assess fatigue and migraine, respectively, and we used consensus epidemiologic guidelines to assess CRS.

This study had several limitations. In general, cross-sectional surveys such as ours cannot assess temporal relationships between exposures and outcomes, and we did not ascertain the onset dates of some symptoms. We note, however, that our UNGD activity metrics could theoretically be used to establish temporality because they can be computed for any date prior to symptom onset. Our ascertainment of self-reported outcomes was susceptible to various types of information bias. For example, despite the fact that our questionnaire did not mention UNGD, individuals residing near UNGD may have over-reported symptoms. There was some evidence of selection bias because survey participants had poorer health (measured by the Charlson comorbidity index) than nonresponders. However, differences in health status were similar across levels of UNGD activity. Another limitation is that our estimates of well development phase durations, although based on published average values, may have been incorrect for individual wells. Further exposure misclassification could have occurred because our UNGD activity metric was based on residential addresses. Participants' exposure to UNGD activity could have been affected by unmeasured factors such as occupation, travel, and time spent outdoors. Additionally, our UNGD activity metric did not allow identification of specific exposures or exposure pathways.

Conclusions

UNGD was associated with CRS, migraine headache, and fatigue symptoms in a large population-based survey. Associations were stronger in patients with two or more outcomes. Our work has several advantages over previous studies, making it an important addition to the growing body of evidence that UNGD is associated with adverse health effects. Further research, including more sophisticated exposure and outcome measurements, is necessary to evaluate whether these associations are causal and to elucidate the mechanisms for these findings.

References

Adgate JL, Goldstein BD, McKenzie LM. 2014. Potential public health hazards, exposures and health effects from unconventional natural gas development. Environ Sci Technol 48(15):8307–8320.

Bell IR, Baldwin CM, Schwartz GE. 1998. Illness from low levels of environmental chemicals: relevance to chronic fatigue syndrome and fibromyalgia. Am J Med 105(3A):74S–82S.

Bhattacharyya N. 2009. Contemporary assessment of the disease burden of sinusitis. Am J Rhinol Allergy 23(4):392–395.

Burch RC, Loder S, Loder E, Smitherman TA. 2015. The prevalence and burden of migraine and severe headache in the United States: updated statistics from government health surveillance studies. Headache 55(1):21–34.

Casey JA, Curriero FC, Cosgrove SE, Nachman KE, Schwartz BS. 2013. High-density livestock operations, crop field application of manure, and risk of community-associated methicillin-resistant *Staphylococcus aureus* infection in Pennsylvania. JAMA Intern Med 173(21):1980–1990.

Casey JA, Savitz DA, Rasmussen SG, Ogburn EL, Pollak J, Mercer DG, et al. 2016. Unconventional natural gas development and birth outcomes in Pennsylvania, USA. Epidemiology 27(2):163–172.

Charlson ME, Pompei P, Ales KL, MacKenzie CR. 1987. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 40(5):373–383.

Drollette BD, Hoelzer K, Warner NR, Darrah TH, Karatum O, O'Connor MP, et al. 2015. Elevated levels of diesel range organic compounds in groundwater near Marcellus gas operations are derived from surface activities. Proc Natl Acad Sci U S A 112(43):13184–13189.

EIA (U.S. Energy Information Administration). 2015. Natural Gas Gross Withdrawals and Production. http://www.eia.gov/dnav/ng/ng_prod_sum_dcu_NUS_a.htm [accessed 3 November 2015].

Fokkens WJ, Lund VJ, Mullol J, Bachert C, Alobid I, Baroody F, et al. 2012. EPOS 2012: European position paper on rhinosinusitis and nasal polyps 2012. A summary for otorhinolaryngologists. 2012. Rhinology 50(1):1–12.

Friedman DI, De ver Dye T. 2009. Migraine and the environment. Headache 49(6):941–952.

Gaines M, Ziegler T. 2013. Shale Energy Development's Effect on the Posting, Bonding, and Maintenance of Roads in Rural PA. http://extension.psu.edu/natural-resources/natural-gas/webinars/shale-energy-developments-effect-on-the-posting-bonding-and-maintenance-of-roads-in-rural-pa [accessed 30 October 2015].

Griffith JP, Zarrouf FA. 2008. A systematic review of chronic fatigue syndrome: don't assume it's depression. Prim Care Companion J Clin Psychiatry 10(2):120–128.

Hastan D, Fokkens WJ, Bachert C, Newson RB, Bislimovska J, Bockelbrink A, et al. 2011. Chronic rhinosinusitis in Europe—an underestimated disease. A GA^{2}LEN study. Allergy 66(9):1216–1223.

Jackson RB, Vengosh A, Darrah TH, Warner NR, Down A, Poreda RJ, et al. 2013. Increased stray gas abundance in a subset of drinking water wells near Marcellus shale gas extraction. Proc Natl Acad Sci U S A 110(28):11250–11255.

Kassotis CD, Tillitt DE, Davis JW, Hormann AM, Nagel SC. 2014. Estrogen and androgen receptor activities of hydraulic fracturing chemicals and surface and ground water in a drilling-dense region. Endocrinology 155(3):897–907.

Kovats S, Depledge M, Haines A, Fleming LE, Wilkinson P, Shonkoff SB, et al. 2014. The health implications of fracking. Lancet 383(9919):757–758.

Lipsitch M, Tchetgen Tchetgen E, Cohen T. 2010. Negative controls: a tool for detecting confounding and bias in observational studies. Epidemiology 21(3):383–388.

Lipton RB, Bigal ME, Diamond M, Freitag F, Reed ML, Stewart WF, et al. 2007. Migraine prevalence, disease burden, and the need for preventive therapy. Neurology 68(5):343–349.

Lipton RB, Dodick D, Sadovsky R, Kolodner K, Endicott J, Hettiarachchi J, et al. 2003. A self-administered screener for migraine in primary care: the ID Migraine validation study. Neurology 61(3):375–382.

Liu AY, Curriero FC, Glass TA, Stewart WF, Schwartz BS. 2012. Associations of the burden of coal abandoned mine lands with three dimensions of community context in Pennsylvania. ISRN Public Health 2012:251201, doi: 10.5402/2012/251201.

Macey GP, Breech R, Chernaik M, Cox C, Larson D, Thomas D, et al. 2014. Air concentrations of volatile compounds near oil and gas production: a community-based exploratory study. Environ Health 13:82, doi: 10.1186/1476-069X-13-82.

Maloney KO, Yoxtheimer DA. 2012. Production and disposal of waste materials from gas and oil extraction from the Marcellus Shale play in Pennsylvania. Environ Pract 14(04):278–287.

Mitka M. 2012. Rigorous evidence slim for determining health risks from natural gas fracking. JAMA 307(20):2135–2136.

New York State Department of Environmental Conservation. 2015. *Final Supplemental Generic Environmental Impact Statement on the Oil, Gas, and Solution Mining Regulatory Program*. http://www.dec.ny.gov/docs/materials_minerals_pdf/fsgeis2015.pdf [accessed 10 October 2015].

Paulik LB, Donald CE, Smith BW, Tidwell LG, Hobbie KA, Kincl L, et al. 2015. Impact of natural gas extraction on PAH levels in ambient air. Environ Sci Technol 49(8):5203–5210.

Pennsylvania Department of Environmental Protection. 2016. PA DEP Oil & Gas Reporting Website - Welcome. https://www.paoilandgasreporting.state.pa.us/publicreports/Modules/Welcome/Welcome.aspx [accessed 19 December 2016].

Potter F. 1988. Survey of procedures to control extreme sampling weights. Proceeding of the Survey Research Methods Section of the American Statistical Association. http://www.websm.org/uploadi/editor/1368363852Potter_1988_Survey_of_procedures_to_control_extreme_sampling_weights.pdf [accessed 19 December 2016].

Powers M, Saberi P, Pepino R, Strupp E, Bugos E, Cannuscio CC. 2015. Popular epidemiology and "fracking": citizens' concerns regarding the economic, environmental, health and social impacts of unconventional natural gas drilling operations. J Community Health 40(3):534–541.

Rabinowitz PM, Slizovskiy IB, Lamers V, Trufan SJ, Holford TR, Dziura JD, et al. 2015. Proximity to natural gas wells and reported health status: results of a household survey in Washington County, Pennsylvania. Environ Health Perspect 123:21–26, doi: 10.1289/ehp.1307732.

Ranjith G. 2005. Epidemiology of chronic fatigue syndrome. Occup Med (Lond) 55(1):13–19.

Ricci JA, Chee E, Lorandeau AL, Berger J. 2007. Fatigue in the U.S. workforce: prevalence and implications for lost productive work time. J Occup Environ Med 49(1):1–10.

Saberi P, Propert KJ, Powers M, Emmett E, Green-McKenzie J. 2014. Field survey of health perception and complaints of Pennsylvania residents in the Marcellus Shale region. Int J Environ Res Public Health 11(6):6517–6527.

Sauro KM, Becker WJ. 2009. The stress and migraine interaction. Headache 49(9):1378–1386.

Shashy RG, Moore EJ, Weaver A. 2004. Prevalence of the chronic sinusitis diagnosis in Olmsted County, Minnesota. Arch Otolaryngol Head Neck Surg 130(3):320–323.

Sjöstrand C, Savic I, Laudon-Meyer E, Hillert L, Lodin K, Waldenlind E. 2010. Migraine and olfactory stimuli. Curr Pain Headache Rep 14(3):244–251.

Steinzor N, Subra W, Sumi L. 2013. Investigating links between shale gas development and health impacts through a community survey project in Pennsylvania. New Solut 23(1):55–83.

Sundaresan AS, Hirsch AG, Storm M, Tan BK, Kennedy TL, Greene JS, et al. 2015. Occupational and environmental risk factors for chronic rhinosinusitis: a systematic review. Int Forum Allergy Rhinol 5(11):996–1003, doi: 10.1002/alr.21573.

Tan BK, Kern RC, Schleimer RP, Schwartz BS. 2013. Chronic rhinosinusitis: the unrecognized epidemic. Am J Respir Crit Care Med 188(11):1275–1277.

Townsend P. 1987. Deprivation. J Soc Policy 16(2):125–146.

U.S. Census Bureau. (2010). TIGER/LineShapefiles (machine- readable datafiles). http://www.census.gov/geo/maps-data/data/tiger-line.html [accessed 24 December 2016].

Werner AK, Vink S, Watt K, Jagals P. 2015. Environmental health impacts of unconventional natural gas development: a review of the current strength of evidence. Sci Total Environ 505:1127–1141.

ATTACHMENT C

STUDY 15

PLOS ONE

RESEARCH ARTICLE

Unconventional Gas and Oil Drilling Is Associated with Increased Hospital Utilization Rates

Thomas Jemielita[1☯], George L. Gerton[2☯], Matthew Neidell[3], Steven Chillrud[4], Beizhan Yan[4], Martin Stute[4], Marilyn Howarth[2], Pouné Saberi[2], Nicholas Fausti[2], Trevor M. Penning[2], Jason Roy[1], Kathleen J. Propert[1], Reynold A. Panettieri, Jr.[2]*

1 Department of Biostatistics, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, United States of America, **2** Center of Excellence in Environmental Toxicology (CEET), Airways Biology Initiative, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, United States of America, **3** Department of Health Policy and Management, Mailman School of Public Health, Columbia University, New York, New York, United States of America, **4** Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York, United States of America

☯ These authors contributed equally to this work.
* rap@mail.med.upenn.edu

OPEN ACCESS

Citation: Jemielita T, Gerton GL, Neidell M, Chillrud S, Yan B, Stute M, et al. (2015) Unconventional Gas and Oil Drilling Is Associated with Increased Hospital Utilization Rates. PLoS ONE 10(7): e0131093. doi:10.1371/journal.pone.0131093

Editor: Jaymie Meliker, Stony Brook University, Graduate Program in Public Health, UNITED STATES

Received: February 2, 2015

Accepted: May 28, 2015

Published: July 15, 2015

Data Availability Statement: All relevant data are available in the paper and its Supporting Information files.

Funding: This work was funded by grants from the National Institute of Environmental Health Sciences (www.niehs.nih.gov): P30-ES013508 (GLG MH PS NF TMP JR KJP RAP) and P30-ES009089 (GLG MN SC BY MS MH PS NF TMP JR KJP RAP). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Over the past ten years, unconventional gas and oil drilling (UGOD) has markedly expanded in the United States. Despite substantial increases in well drilling, the health consequences of UGOD toxicant exposure remain unclear. This study examines an association between wells and healthcare use by zip code from 2007 to 2011 in Pennsylvania. Inpatient discharge databases from the Pennsylvania Healthcare Cost Containment Council were correlated with active wells by zip code in three counties in Pennsylvania. For overall inpatient prevalence rates and 25 specific medical categories, the association of inpatient prevalence rates with number of wells per zip code and, separately, with wells per km^2 (separated into quantiles and defined as well density) were estimated using fixed-effects Poisson models. To account for multiple comparisons, a Bonferroni correction with associations of $p<0.00096$ was considered statistically significant. Cardiology inpatient prevalence rates were significantly associated with number of wells per zip code ($p<0.00096$) and wells per km^2 ($p<0.00096$) while neurology inpatient prevalence rates were significantly associated with wells per km^2 ($p<0.00096$). Furthermore, evidence also supported an association between well density and inpatient prevalence rates for the medical categories of dermatology, neurology, oncology, and urology. These data suggest that UGOD wells, which dramatically increased in the past decade, were associated with increased inpatient prevalence rates within specific medical categories in Pennsylvania. Further studies are necessary to address healthcare costs of UGOD and determine whether specific toxicants or combinations are associated with organ-specific responses.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

The United States now leads the world in producing natural gas from shale formations. Shale gas accounted for 40% of all natural gas produced in 2012 [1–4]. In comparison to the early 2000s, natural gas production in the US has increased with more than a 30% increase in production, due in part to the cost-effective combination of horizontal drilling and hydraulic fracturing [1–4].

Unconventional gas and oil drilling (UGOD), including hydraulic fracturing or "fracking", refers to all activities that extract natural gas and oil from rock formations. At distances from 1 to 2 miles below the earth's surface, tight rock formations impede natural gas and oil flow into a drill-hole [3]. Common reservoirs that contain natural gas and oils include: porous sandstones, limestones, dolomite rocks, shale rocks, and coal beds. Hydraulic fracturing and horizontal drilling methods can effectively extract these resources. Typically, after drilling is complete, fissures are formed using a perforating gun; a mixture of water, proppants and hydraulic fracturing chemicals is then pumped into the rock [3,5]. Consequently, the fissures remain open to liberate the gas. These substances as well as contaminants released from the shale are present in the flowback water. Contaminants include naturally occurring radioactive materials [3,4], toxic organics and metals that may enter ground water, contaminating water supplies especially if leakage occurs from casement failure or from holding ponds for waste water [6,7]. Other toxicants and volatile organic compounds, such as benzene, ethylbenzene, toluene and xylene or radionuclides, have been seen in ground waters impacted by UGOD spills [8] or surface waters receiving UGOD-related waste water [9]. The general lack of published baseline (i.e., pre-UGOD) data has limited efforts to associate contamination in drinking water wells to UGOD activities [10]. Additionally, exhaust produced by diesel trucks and off-site diesel engines, as well as emissions from other UGOD activities (e.g., venting, flaring, compressor stations, etc.) may also affect local air quality with potential impact on health [11–13]. Plausibly, increased noise pollution, truck traffic, and psychosocial stress due to community change, which occur due to increased hydro-fracking activity, could impact public health [11].

Despite the growth in hydraulic fracturing, the health consequences of UGOD are unclear [3,4,14,15]. In Pennsylvania (PA), a rise in hydraulic fracturing has raised health concerns, especially since the Marcellus Shale formation underlies two-thirds of Pennsylvania [16]. In northeastern Pennsylvania, most wells were drilled for dry gas rather than gas and oil [17]. We postulate that increases in active or producing wells in Pennsylvania from 2007 to 2011 are associated with increases in inpatient prevalence rates. Three counties, which lie on the Marcellus Shale formation along the northern border of PA, were chosen for this study: Bradford, Susquehanna, and Wayne. Importantly, zip codes in Bradford and Susquehanna Counties significantly increased UGOD over this time period. These counties are some of the greatest producers of natural gas in Pennsylvania, generating 489 million cubic feet of natural gas from 598 wells in 2011 [18]. In contrast, zip codes in Wayne County have no active wells [18]. Specifically, we evaluated the association between inpatient prevalence rates and well density within 25 different medical categories, as well as overall inpatient prevalence rates.

Materials and Methods

This study is an ecological study with the goal of assessing the association between hydro-fracking activity and health care use. Zip code specific inpatient counts were obtained from the time frame of 2007–2011. Only zip codes from the counties Bradford, Susquehanna, and Wayne were considered. For our analysis, only inpatient records for people who resided in one of these three counties were included. Inpatient records of people who came to a hospital in these counties, but did not reside in one of these counties, were excluded. These counties were of particular interest, since Wayne had no hydro-fracking activity between 2007 and 2011, while Bradford and Susquehanna saw increased hydro-fracking activity. Inpatient counts were then converted into inpatient prevalence rates (details in Statistical Methods). Furthermore, for each zip code, we obtained the number of wells for each year in 2007–2011. In total, there were 67 zip codes considered, with five inpatient prevalence rates/well counts each. Inpatient prevalence rates were the primary outcome of interest with wells as the primary predicator of interest.

Health Utilization Data

Truven Health Analytics (THA) purchased UB92/UB04 inpatient discharge datasets from the Pennsylvania Health Care Cost Containment Council (PHC4). The PHC4 datasets contain all inpatient hospital discharge records, including those for psychiatric and/or behavioral health, rehabilitation, and drug and alcohol treatment, for patients hospitalized in Pennsylvania. Skilled nursing facility (SNF), swing bed, transitional care unit, 23-hour observation, and hospice records are not included. After receipt of state discharge datasets, THA decoded supplied values, checked the validity of information submitted and standardized the format. The ICD-9 diagnosis codes and MSDRGs included in the data pulls can be found in S1 Table, in the supplemental material section.

Truven Health pulled discharge records for patients residing in any of the Bradford, Susquehanna, and Wayne County zip codes for calendar years 2007, 2008, 2009, 2010, and 2011. Treatment records for those patients hospitalized outside of Pennsylvania were not captured. In addition, THA excluded patient records for those patients with dentistry, HIV, and neurosurgery DRGs.

Insurance Coverage Estimates (ICE) Overview

ICE reports by THA showed the total number of people covered by seven different types of insurance by zip code, age group, and sex for every market in the United States. The seven different types of insurance are Medicaid, Medicare, dual eligible, private employer sponsored, private exchanges, private direct, and uninsured. Every person in a zip code who is a resident is assigned an insurance category based on his or her primary insurance coverage. Only non-residents of zip codes were excluded from the analysis.

Demographics Methodology

THA acquires all of its demographic data from The Nielsen Company statistics for every zip code in the United States. Nielsen bases their estimates on products of the United States Census Bureau, including the 2010 Census Summary File 1 (SF1). Details of the methodology and definitions used to create the SF1 data, including field definitions and the 2010 Census questionnaire, are available in the Census 2010 Data Definitions publication [19].

Mapping of Unconventional Gas Wells in Bradford and Susquehanna Counties in Pennsylvania

To create maps of the unconventional gas well locations, the complete data set for 2000–2013 was downloaded as comma separated values (CSV) from the Pennsylvania Department of Environmental Protection Oil and Gas Reporting Website [20] and imported into FileMaker Pro Advanced 13.0.v.3 for further processing. For Fig 1, the data were filtered for unconventional, drilled wells that produced gas in the noted year. We use the state's categorization, such that: "An unconventional gas well is a well that is drilled into an unconventional formation, which is defined as a geologic shale formation below the base of the Elk Sandstone or its

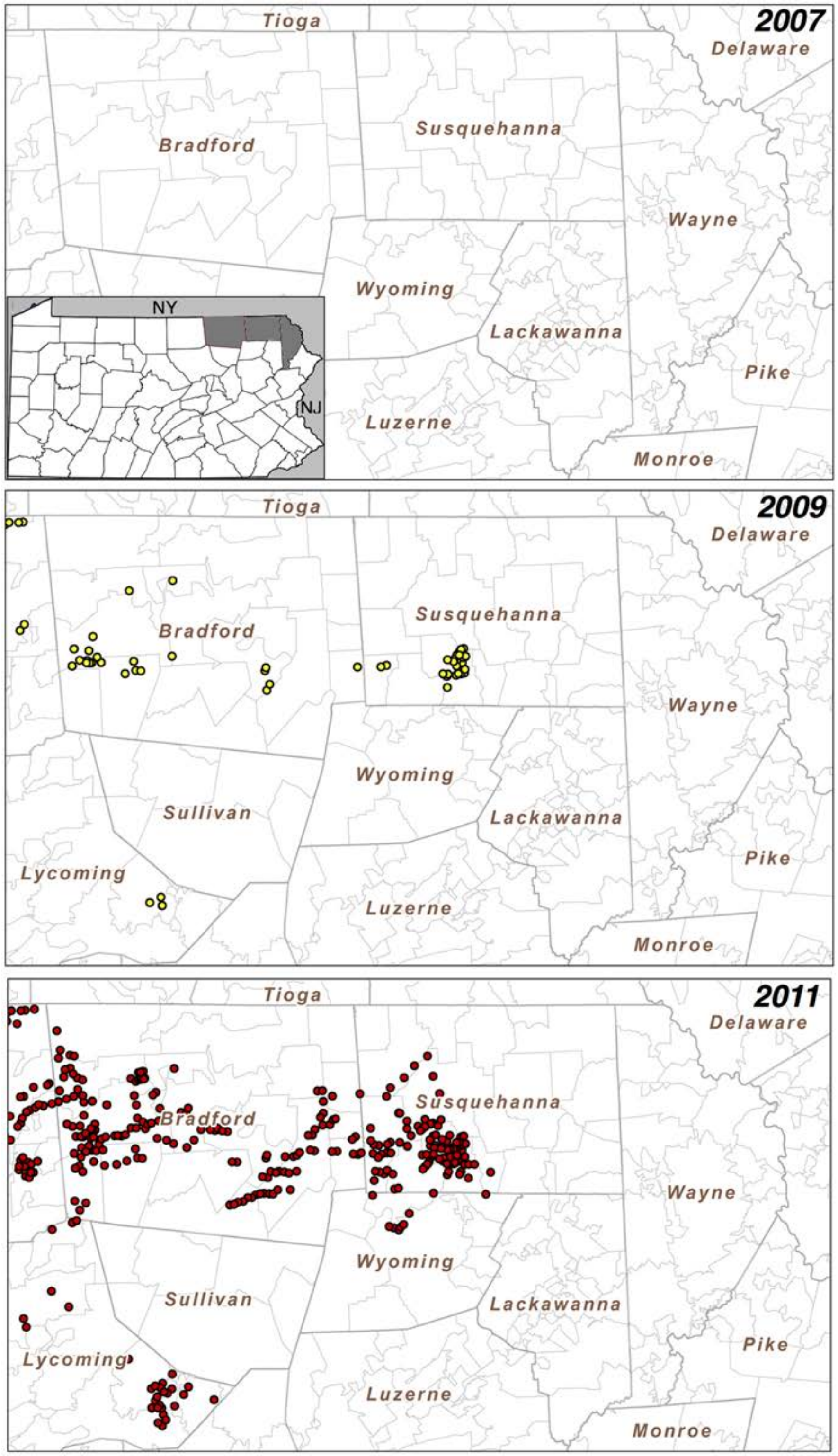


Fig 1. Pennsylvania active wells over time. Pennsylvania active wells in Bradford and Susquehanna Counties increased markedly from 2007 to 2011. Wells are shown as colored dots. From 2007 to 2011, Wayne County effectively had no active wells. Insert in the first panel shows location of Bradford, Susquehanna and Wayne Counties within Pennsylvania.

doi:10.1371/journal.pone.0131093.g001

geologic equivalent where natural gas generally cannot be produced except by horizontal or vertical well bores stimulated by hydraulic fracturing." These data were exported as a DBF file and imported into ESRI ArcGIS v.10.2 to map the locations of the producing wells. In any given year, only wells that produced gas in that year are shown in Fig 1. For example, if a well produced gas in 2007 but did not in 2011, then this well would only appear on the 2007, but not on the 2011 map.

Statistical Analysis

Statistical analysis was performed using STATA 13 software (StataCorp LP, College Station, Texas). Our data included the number of wells and inpatient counts for all combinations of year, medical category (25 total), and zip code within the three chosen counties in PA. In total, after excluding eight zip codes that had no available population information, 67 zip codes were considered. Only inpatient counts for patients that resided in one of three counties were considered. For each zip code, population and total area per square kilometer (km) data were obtained from the US Census 2010. Importantly, zip code specific population and total area per square km were the same for each year in 2007–2011. Number of wells is defined as the number of wells within a specific zip code for a certain year. All data are generated from active wells. We assume that once a well is active in 2007, this same well remains active for the time frame of 2007–2011. For example, if there are 3 wells in 2007 and 8 wells in 2008 for some zip code, then we assume that there were an additional 5 wells created between 2007 and 2008. This is in contrast to the definition of active wells for the mapping, where a well can move from being active to inactive in any given year in 2007–2011. Given the 5-year observation period, very few active wells became inactive. In addition, the actual date of inactivity could not be accurately defined. Furthermore, it is possible that once a well becomes inactive, it could still impact the surrounding community for some period of time. Thus, for the statistical analysis, once an active well enters at any given year, we assume the well remains active for the remainder of the years. In addition to the count of wells, we also generated wells per square km (wells/km^2), which is the number of wells divided by the total area per square km (at the zip code level); we defined this variable as well density. We analyzed both exposure variables (count and density) because, a priori, it was unclear whether the number of wells or the density of wells would have a stronger association with health outcomes. Zip code specific inpatient prevalence rates for each medical category (and overall) were calculated by dividing the zip code specific number of inpatient counts per year by the population of the zip code. The inpatient prevalence rates were then converted into prevalence rates per year per 100 people and treated as the primary outcome for modeling. We now refer to prevalence rates per year per 100 people when we discuss inpatient prevalence rates.

Since we examined a relatively brief interval of time (2007–2011), we postulated that in a given zip code, inpatient prevalence rates would be relatively stable. Our goal was to obtain an un-confounded estimate of the association between inpatient prevalence rates and wells. However, it is possible that observable or unobservable zip code characteristics will be correlated with wells and inpatient prevalence rates. Accordingly, we used conditional fixed effects Poisson regression, where the fixed effects are the zip codes. This controls for all possible characteristics of the zip codes, both measured and unmeasured, that did not change during the period of observation. Thus, if zip codes that consistently have high rates of inpatient prevalence rates are more likely to have more wells over time, this will be accounted for in the model. Alternatively, if there are zip code-level changes from 2007–2011 that affect the number of wells and inpatient prevalence rates, this model will not account for this. Essentially, our methodology captures the association between and within zip code changes in wells and inpatient prevalence rates. Furthermore, to account for potential over-dispersion, we use robust standard errors [21]. These robust standard errors are cluster-robust estimates, where the clusters are the individual zip codes in this case. Two sets of analyses are then done to investigate the relationship between inpatient prevalence rates and wells.

The first set of analyses relates inpatient prevalence rates to number of wells. Exploratory analyses suggested that the relationship between the log of the inpatient prevalence rates (Poisson model uses a log link) and number of wells was linear. Thus, for these analyses, prediction variables were the number of wells and year (2007–2011). This assumes a linear relationship between number of wells and inpatient prevalence rates, as well as a linear association between inpatient prevalence rates and year. Note that the primary predictor of interest was the number of wells. This will be referred to as the ***number of wells analysis***.

Furthermore, while exploratory analyses suggested a linear relationship between the log of inpatient prevalence rates and number of wells, we also reasoned that a quadratic relationship between the log of inpatient prevalence rates and number of wells was plausible. Subsequently, we also examined whether there exists a non-linear relationship between number of wells and inpatient prevalence rates. Accordingly, a second model incorporated a quadratic relationship between number of wells and inpatient prevalence rates, for each medical category and overall. Prediction variables within this model were year (2007–2011)/wells, and wells^2.

The second set of analyses relates inpatient prevalence rates to wells/km^2 (well density). However, the relationship between inpatient prevalence rates and well density is highly non-linear and heavily influenced by observations that have extremely high wells/km^2. For example, one zip code located in Bradford had 16.9 wells/km^2 and 23.4 wells/km^2 in 2010 and 2011, respectively, while 99% of all wells/km^2 observations had fewer than 4.28 wells/km^2. Subsequently, we opted to separate wells/km^2 into four levels based on quantiles as shown in Table 1. We set Q0wells to be the reference category and all the other levels (Q1wells, Q2wells, Q3wells) to have separate dummy variables. ***This will be referred to as the quantile analysis***.

Our analysis investigates the association of increasing wells/km^2 on inpatient prevalence rates, while allowing for separate associations depending on the magnitude of well/km^2. We, however, recognize that by using quantiles, we lose information and cannot make inference on explicit changes in well density. Furthermore, while our cut-offs are somewhat arbitrary, the goal is to determine whether increased well density is positively associated with inpatient prevalence rates, which is accomplished by this modeling approach. Overall, the primary predictors for this set of analyses included Q1wells, Q2wells, Q3wells, and year. We test the overall Wald test that the coefficients Q1wells = Q2wells = Q3wells = 0.

For all analyses, risk ratios were obtained by taking the exponential of the regression coefficient estimates. Year is recoded into 2007 = 0, 2008 = 1, 2009 = 2, 2010 = 3, and 2011 = 4. We model each medical category separately as well as the overall inpatient prevalence rates, for a total of 26 models per set of analyses. Furthermore, to adjust for multiple comparisons, we use a Bonferroni correction to adjust for testing 25 different medical categories and overall inpatient prevalence rates in both sets of analyses (52 tests). Using an initial level of significance of

Table 1. Definition of quantiles by wells/km^2.

	Q0wells	Q1wells	Q2wells	Q3wells
wells/km^2	0	(0, 0.168]	(0.168, 0.786]	>0.786
Quantile	(0, 65.97]	(65.97, 80]	(80, 90.15]	(90.15, 100]

Note: (A, B] indicates that A is excluded from the range, and B is included.

doi:10.1371/journal.pone.0131093.t001

0.05, this means we reject the null hypothesis that wells are not associated with hospitalizations for $p<0.00096$.

Sensitivity analyses were also performed to determine if removing a specific zip code with much higher inpatient prevalence rates or with much higher well density affected inference. Thus, we removed the specific zip code(s) and recalculated the conditional fixed effects Poisson models, checking to see if the general inference changed.

All of the data obtained for this study were received anonymized and de-identified from Truven Health Analytics. The data were provided as summary information, and there were no unique identifiers. The University of Pennsylvania Committee on the Study of Human Subjects deemed this work non-human subject research.

Results

Subject Demographics by County

The three Pennsylvania counties chosen for analysis were Bradford, Susquehanna, and Wayne. These counties were selected given the completeness of health care utilization data from 2007 to 2011. Bradford and Susquehanna Counties also had large increases in active wells over this time period. Wayne County, which effectively had no active wells from 2007 to 2011, served as a unique control population whose demographics were comparable to Bradford and Susquehanna Counties. The total number of residents as per the most recent census in Bradford, Susquehanna, and Wayne Counties was 157,311. As shown in Table 2, the summary of subject demographics for the three Pennsylvania counties obtained from US census data was comparable. Even though the statistical analysis is done at the zip code level, a county level demographic table is an informative summary of the zip codes that are within the counties. Each county is

Table 2. Characteristics Table for PA Counties.

		Bradford	Susquehanna	Wayne
Population		62,622	43,356	51,548
Overall Hospitalizations 2007–2011		39,821	22,559	30,425
Age (median)		43.4	45.1	45.9
Male %		49.5	50.4	52.8
High School Graduate, percent of person age 25+ %		86.6	88.1	87.4
Bachelor Degree or Higher, percent of person age 25+ %		16.4	16.1	18.4
Median Income (2008–2012) $		44,650	46,815	50,153
Race %	White	97.4	98.0	94.7
	Black	0.6	0.4	3.5
	Asian	0.6	0.3	0.5
	Other	1.4	1.3	1.3
Median Number of Wells	2007	0	0	0
	2008	1	0	0
	2009	13	0	0
	2010	81	1	0
	2011	149	6	0
Number of Zip Codes with >0 Wells (%)	2007	4 (19)	2 (9)	0 (0)
	2008	12 (57)	4 (17)	0 (0)
	2009	16 (76)	8 (35)	0 (0)
	2010	20 (95)	12 (52)	0 (0)
	2011	20 (95)	16 (70)	0 (0)

doi:10.1371/journal.pone.0131093.t002

one data point, so no formal statistical comparison is possible. There were no striking differences among the three counties. The subjects were predominantly Caucasian with few people obtaining higher than a high school diploma. Further, the median income was similar among the counties. Table 2 also illustrates the growth in hydro-fracking activity from 2007 to 2011 for Bradford and Susquehanna. By 2011, 95% of the zip codes in Bradford had at least one well, while 70% of the zip codes in Susquehanna had at least one well.

Inpatient Prevalence Rates by Medical Category

Table 3 shows the median inpatient prevalence rates and median inpatient counts, along with the interquartile range (IQR), for each medical category as well as overall. The median inpatient prevalence rates and median inpatient counts are to be interpreted at the zip code level. Notably, there are a number of categories with very low (or zero) median inpatient prevalence rates and median inpatient counts. Furthermore, cardiology inpatient prevalence rates/inpatient counts seem to be higher than the other medical categories (excluding overall), with a median cardiology inpatient prevalence rate of 1.99 and a median cardiology inpatient count of 18.

Table 3. Median Inpatient Prevalence Rates per 100 people and Median Inpatient Counts, by Medical Category.

Medical Category	Median Inpatient Prevalence Rate (IQR)	Median Inpatient Counts (IQR)
Inpatient total	12.12 (10.05, 14.84)	106 (41, 272)
Cardiology	1.99 (1.42, 2.56)	18 (6, 46)
Dermatology	0.21 (0.09, 0.34)	2 (1, 6)
Endocrine	0.22 (0.01, 0.37)	2 (0.5, 7)
Gastroenterology	1.02 (0.71, 1.43)	10 (3, 27)
General medicine	0.58 (0.32, 0.88)	5 (2, 14)
Generals surgery	0.75 (0.47, 1.01)	6 (3, 19)
Gynecology	0.14 (0, 0.26)	2 (0, 5)
Hematology	0.05 (0, 0.14)	1 (0, 3)
Neonatology	0.12 (0, 0.23)	2 (0, 4)
Nephrology	0.34 (0.18, 0.53)	3 (1, 9)
Neurology	0.58 (0.35, 0.88)	5 (2, 16)
Normal newborns	0.68 (0.41, 0.99)	6 (2, 17)
Ob/delivery	0.84 (0.52, 1.12)	7 (2.5, 21)
Oncology	0.17 (0, 0.29)	2 (0, 6)
Ophthalmology	0 (0, 0)	0 (0, 0)
Orthopedics	1.08 (0.72, 1.42)	10 (4, 26)
Other/ob	0 (0, 0.09)	0 (0, 2)
Otolaryngology	0.08 (0, 0.17)	1 (0, 3)
Psych/drug abuse	0.52 (0.27, 0.85)	5 (2, 16)
Pulmonary	1.18 (0.84, 1.69)	10 (4, 28)
Rheumatology	0 (0, 0.09)	0 (0, 2)
thoracic surgery	0.08 (0, 0.16)	1 (0, 3)
Trauma	0.03 (0, 0.09)	1 (0, 2)
Urology	0.17 (0, 0.27)	2 (0, 5)
Vascular surgery	0.09 (0, 0.19)	1 (0, 3)

Note: Median inpatient prevalence rates/median inpatient counts for each medical category and overall are presented, along with the interquartile range (IQR). Median inpatient prevalence rates/median inpatient counts are interpreted at the zip code level.

doi:10.1371/journal.pone.0131093.t003

Geographic Location of Wells from 2007 to 2011

Given the demand in accessing the Marcellus Shale for UGOD, we next examined the active wells over time. There was a dramatic increase in the number of active wells from 2007 to 2011 as shown in Fig 1. In Bradford and Susquehanna Counties, there were substantial increases in the total numbers of wells with two zip codes having the greatest number of wells with 400 and 395, respectively. In Wayne County, there were no active wells from 2007 to 2011. The most dramatic increases were in Bradford County where wells were acquired more uniformly than those in Susquehanna County, where active wells were primarily located in the southwest corner as shown in Fig 1. Gas production tracked with increasing active well numbers from 2007–2013 as shown in Fig 2. These data suggest that if UGOD continues at the rates observed between 2007 and 2011, well densities are likely to continue to increase. Within the counties, there were also profound differences in wells by zip code. For example, in 2011, 31 zip codes had no wells, but 17 zip codes had at least 100 wells.

Increases in Active Wells Are Associated with Increases in Inpatient Prevalence Rates

Given the rapid increase in wells, we reasoned that increases in wells were associated with changes in inpatient prevalence rates. Of the 67 zip codes examined in the three counties, total inpatient counts from 2007 to 2011 were 92,805. There was marked variation in inpatient prevalence rates across zip codes. Specifically, one zip code had a much higher combined inpatient rate as compared with others as shown in Fig 3. Fig 3 also shows that, within each zip code, the contribution by year was comparable, suggesting that within each zip code, the inpatient rates are relatively stable from 2007–2011 Indeed, the average overall inpatient prevalence rates for 2007–2011 are, respectively, 15.18, 15.30, 14.86, 14.00, 14.25. This indicates that on average, zip code overall inpatient prevalence rates were relatively stable or possibly declining from 2007 to 2011, which mirrors national trends [22]. Fig 4 shows how in 2007, 91% (61/67) of zip codes had no wells. However, by 2011, only 46% (31/67) of zip codes had no wells while 54% of zip codes had least 1 well. Notably, many zip codes had a large number of wells by 2011. 28% (19/67) of zip codes had greater than 0.79 wells/km^2, which equates to 79 wells for every 100 km^2. Importantly, Fig 4 corresponds to the quantile analysis.

To further understand health consequences by disease category, we modeled the 25 top specific medical categories and total inpatients, investigating the association between number of wells and inpatient prevalence rates and the association between well density and inpatient prevalence rates. Only cardiology inpatient prevalence rates were significantly associated with number of wells, taking into account our Bonferroni correction ($p<0.00096$) as shown in Table 4. While other medical categories did not strictly meet the Bonferroni correction boundary, a positive association of well number with inpatient prevalence rates within dermatology, neonatology, neurology, oncology, and urology was also evident. Cardiology and neurology inpatient prevalence rates were also significantly associated with well density as shown in Table 5. Furthermore, these results suggest an almost monotonic increase in the impact of well density on cardiology inpatient prevalence rates, considering how the risk ratio increases moving from quantiles (Q1wells to Q2wells to Q3wells). Evidence also suggests that well density was positively associated within the medical categories of dermatology, endocrine, neurology, oncology, urology, as well as overall inpatient prevalence rates ($p = < 0.05$). Furthermore, for both sets of analyses, the year variable is significantly and negatively associated with inpatient prevalence rates, within the medical categories of gynecology and orthopedics.

In both the number of wells analyses and the well density quantile analyses, cardiology inpatient prevalence rates were significantly associated with wells. Under the quantile analyses,

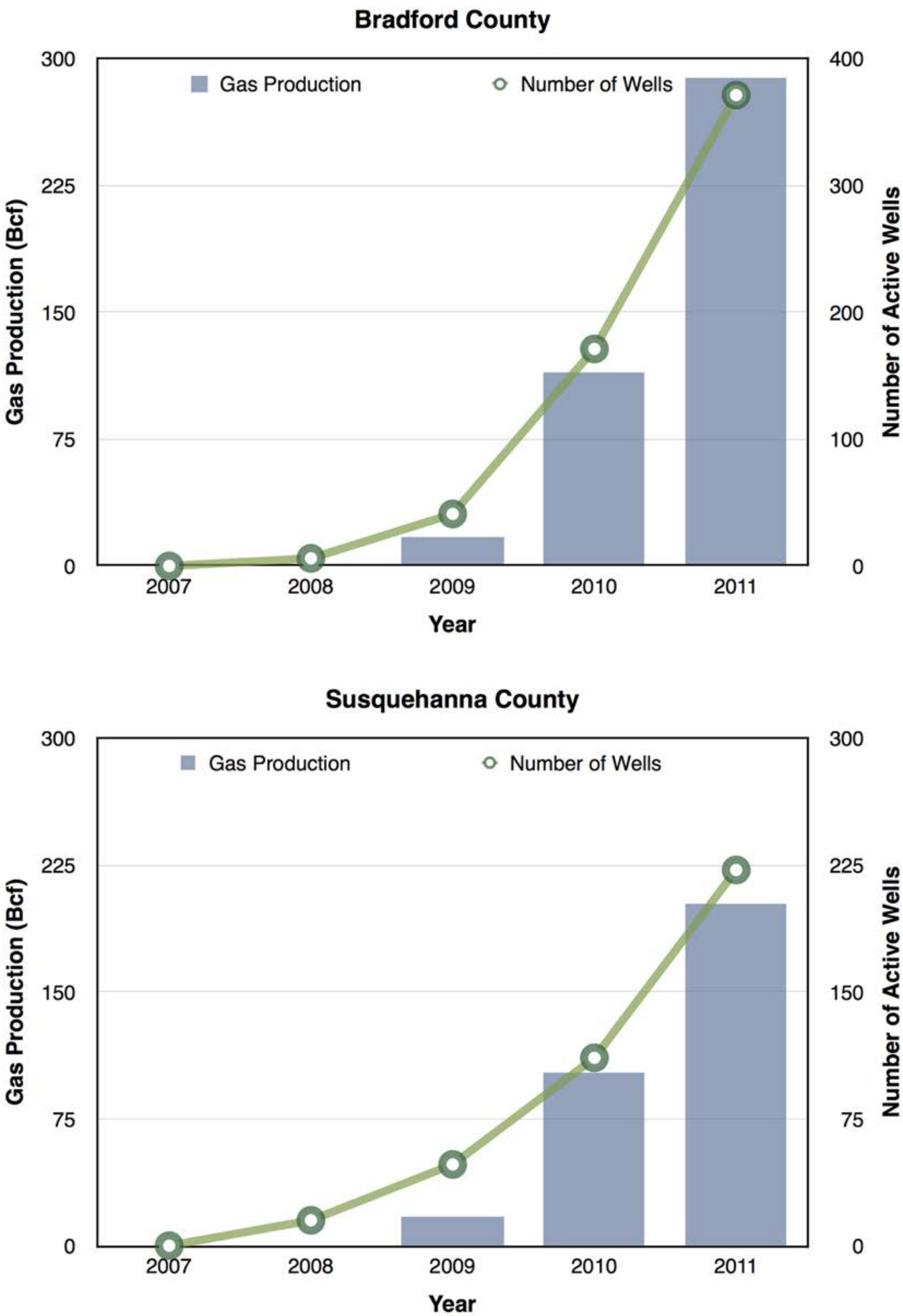


Fig 2. Gas production (histogram) linearly tracked with well number (open circles) from 2007–2011.

doi:10.1371/journal.pone.0131093.g002

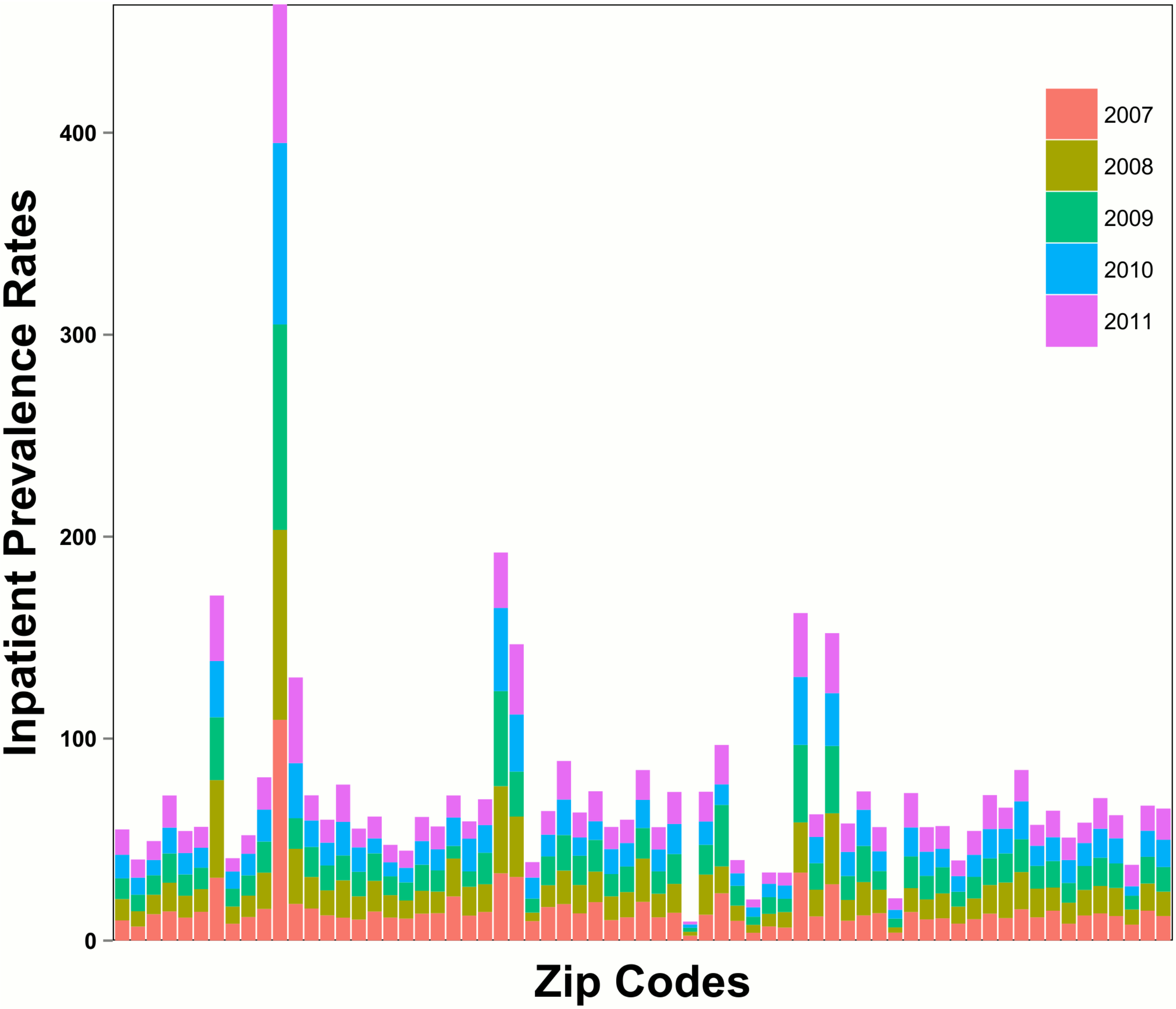


Fig 3. Total inpatient rates by zip code. Total inpatient prevalence rates by zip code. From 2007 to 2011, within a zip code, inpatient prevalence rates are relatively stable.

doi:10.1371/journal.pone.0131093.g003

neurology inpatient prevalence rates were also significantly associated with well density. Also, both sets of analyses show evidence that dermatology, neurology, oncology, and urology inpatient prevalence rates were positively associated with wells. While only the number of wells analyses showed evidence of a positive association between wells and neonatology inpatient prevalence rates, our findings are consistent with other reports suggesting that such illnesses are linked with hydro-fracking [12].

A quadratic association between number of wells and inpatient prevalence rates was also explored. A quadratic relationship seemed to fit the data better than a linear relationship between number of wells and inpatient prevalence rates, within the ophthalmology and neurology categories, where the p-value for the quadratic number of wells term was, respectively, 0.04

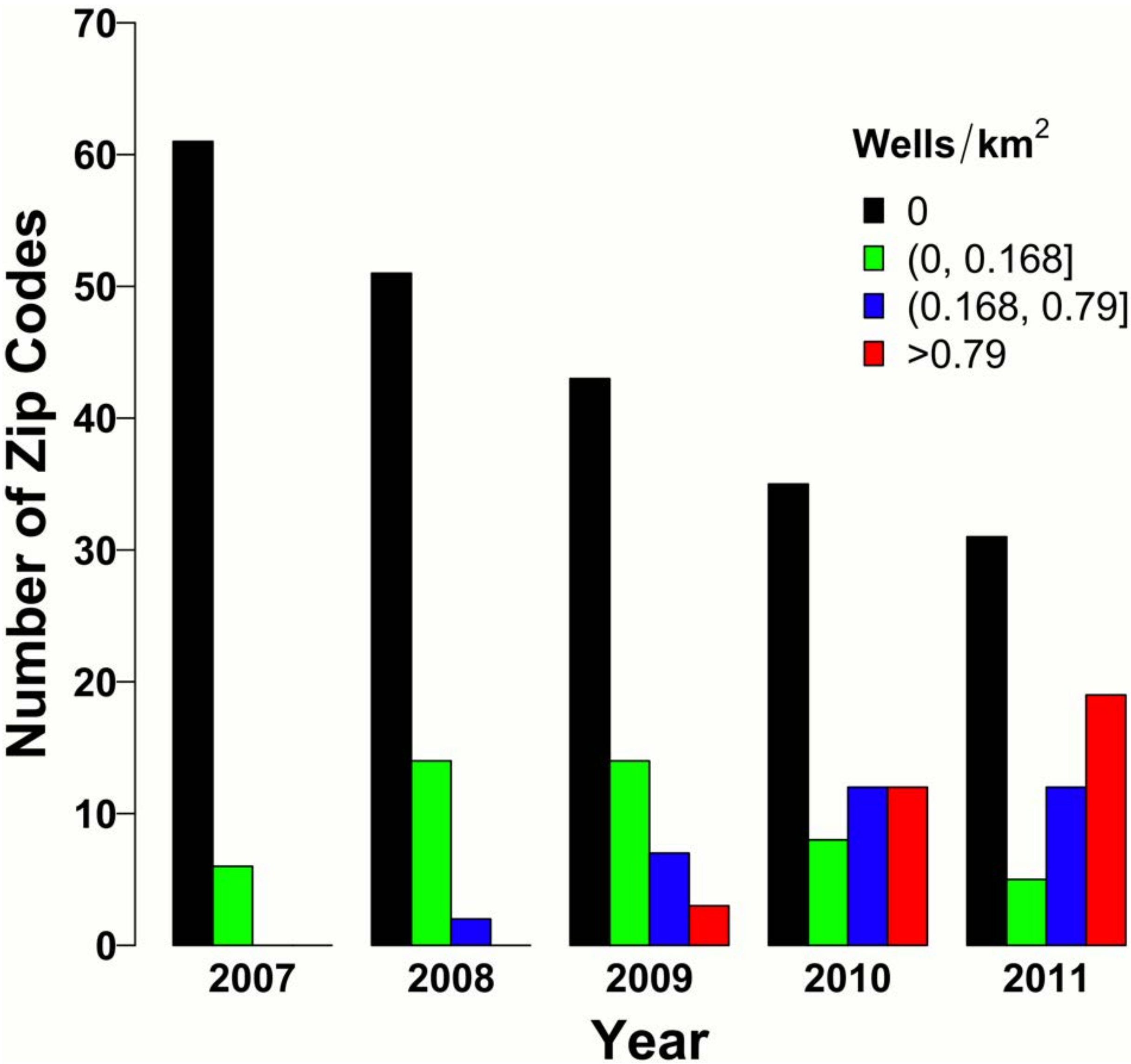


Fig 4. Well density (quantiles) by year. Number of zip codes by well density (quantiles) is presented for each year. In 2007, the majority of zip codes have no wells, but by 2011, the majority of zip codes have at least 1 well.

doi:10.1371/journal.pone.0131093.g004

and 0.004. However, these did not meet the Bonferroni threshold. Furthermore, given Table 3 and the sparsity of ophthalmology inpatient prevalence rates (first three quartiles have no inpatient prevalence rates), it seems unlikely that inference is valid for the ophthalmology models. Given this weak evidence of a quadratic association, results for the quadratic number of wells models are not shown.

In our analysis, one particular zip code had extremely high inpatient prevalence rates compared to other zip codes. Thus, a sensitivity analysis was performed (data not shown). This zip code is located within Wayne County and had no active wells from 2007 to 2011. Removal of this zip code from the analysis had little effect on either the number of wells or the quantile analyses, and there was no change in inference and the estimated risk ratios. Next, a zip code in Bradford had extremely high wells/km^2 in 2010 and 2011, 16.9 wells/km^2 and 23.4 wells/km^2, respectively. Consequently, we explored both sets of analyses without this zip code to

Table 4. Poisson Fixed Effects Models: Number of Wells per Zip Code per Year.

	Wells RR (p-value)	Year RR (p-value)
Inpatient total	1.0003 (0.076)	0.984 (0.128)
Cardiology	**1.0007 (0.0007)**	**0.966 (0.029)**
Dermatology	1.0010 (0.039)	0.977 (0.345)
Endocrine	1.0008 (0.086)	0.963 (0.316)
Gastroenterology	1.0003 (0.338)	0.992 (0.749)
General medicine	1.0002 (0.574)	1.037 (0.022)
Generals surgery	1.0000 (0.849)	1.104 (0.213)
Gynecology	1.0002 (0.708)	0.860 (<0.0001)
Hematology	0.9997 (0.657)	1.023 (0.616)
Neonatology	1.0014 (0.018)	0.959 (0.125)
Nephrology	0.9998 (0.461)	1.025 (0.250)
Neurology	1.0006 (0.037)	1.001 (0.948)
Normal newborns	1.0000 (0.969)	0.963 (0.030)
Ob/delivery	1.0002 (0.411)	0.968 (0.411)
Oncology	1.0015 (0.004)	0.956 (0.081)
Ophthalmology	1.0010 (0.593)	1.084 (0.255)
Orthopedics	0.9993 (0.011)	0.970 (<0.0001)
Other/ob	1.0003 (0.727)	0.899 (0.007)
Otolaryngology	1.0000 (0.982)	0.978 (0.614)
Psych/drug abuse	1.0004 (0.073)	1.035 (0.006)
Pulmonary	1.0000 (0.850)	0.989 (0.482)
Rheumatology	1.0014 (0.043)	0.961 (0.227)
thoracic surgery	1.0011 (0.100)	0.989 (0.708)
Trauma	1.0008 (0.174)	1.021 (0.505)
Urology	1.0010 (0.012)	0.983 (0.464)
Vascular surgery	0.9997 (0.539)	0.948 (0.024)

Note: RR = Risk ratio

doi:10.1371/journal.pone.0131093.t004

determine whether removal of this zip code changed inference. Like the first sensitivity analysis, removal of the Bradford zip code had little effect on inference.

Discussion

We posit that larger numbers of active hydraulic fracturing wells would increase inpatient prevalence rates over time due in part to increases in potential toxicant exposure and stress responses in residents evoked by increases in the hydraulic fracturing work force and diesel engine use. We recognize that a five-year observation period may limit our ability to discern a direct impact on health in the surrounding community but may offer an opportunity to assess hospital utilization rates over time. We examined over 95,000 inpatient records, and thus our study, to our knowledge, represents the most comprehensive one to date to address the health impact of UGOD.

Our data suggests that some but not all medical categories were associated with increases in number of wells, along with increases in well density. Specifically, cardiology inpatient prevalence rates were significantly associated with number of wells and well density, while neurology inpatient prevalence rates were significantly associated with well density. We are struck by the finding that these differences were observable within a short period of time from 2007–2011.

Table 5. Poisson Fixed Effects Models: Quantile Analysis of Wells/km^2.

	Q1 Wells RR (p-value)	Q2 Wells RR (p-value)	Q3 Wells RR (p-value)	Wald Test of all Q Wells = 0	Year RR (p-value)
Inpatient total	0.979 (0.475)	1.069 (0.044)	1.108 (0.041)	P = 0.0058	0.977 (0.013)
Cardiology	**1.021 (0.667)**	**1.142 (0.018)**	**1.27 (0.001)**	**P = 0.0008**	**0.957 (0.004)**
Dermatology	1.051 (0.572)	1.108 (0.429)	1.454 (0.013)	P = 0.0329	0.972 (0.329)
Endocrine	0.975 (0.862)	1.228 (0.045)	1.391 (0.029)	P = 0.0068	0.942 (0.039)
Gastroenterology	0.943 (0.369)	1.12 (0.168)	1.105 (0.364)	P = 0.1101	0.98 (0.406)
General medicine	0.911 (0.234)	0.993 (0.931)	0.985 (0.872)	P = 0.6373	1.037 (0.006)
Generals surgery	0.875 (0.011)	0.921 (0.228)	0.944 (0.424)	P = 0.0669	1.015 (0.157)
Gynecology	0.887 (0.300)	0.938 (0.606)	0.967 (0.849)	P = 0.7549	0.865 (<0.0001)
Hematology	1.202 (0.365)	1.21 (0.320)	1.221 (0.429)	P = 0.7145	0.993 (0.868)
Neonatology	0.994 (0.975)	1.301 (0.152)	1.527 (0.100)	P = 0.0745	0.95 (0.052)
Nephrology	1.115 (0.203)	1.143 (0.227)	1.151 (0.211)	P = 0.5566	1.004 (0.871)
Neurology	**0.922 (0.344)**	**1.157 (0.048)**	**1.188 (0.062)**	**P = 0.0003**	**0.99 (0.542)**
Normal newborns	0.949 (0.481)	0.978 (0.764)	0.964 (0.731)	P = 0.8980	0.965 (0.064)
Ob/delivery	0.958 (0.524)	1.028 (0.670)	1.029 (0.749)	P = 0.4219	0.956 (0.002)
Oncology	1.217 (0.144)	1.415 (0.028)	1.815 (0.002)	P = 0.0166	0.938 (0.022)
Ophthalmology	0.717 (0.381)	1.014 (0.976)	1.116 (0.836)	P = 0.5215	1.099 (0.263)
Orthopedics	0.996 (0.940)	0.981 (0.740)	0.875 (0.130)	P = 0.3591	0.963 (<0.0001)
Other/ob	0.966 (0.885)	1.176 (0.451)	1.264 (0.502)	P = 0.7209	0.879 (0.001)
Otolaryngology	1.052 (0.744)	1.194 (0.412)	1.004 (0.988)	P = 0.5564	0.966 (0.527)
Psych/drug abuse	0.944 (0.307)	0.927 (0.293)	1.13 (0.145)	P = 0.0535	1.039 (0.008)
Pulmonary	1.05 (0.267)	1.097 (0.202)	1.067 (0.572)	P = 0.3050	0.981 (0.306)
Rheumatology	1.091 (0.601)	1.432 (0.159)	1.866 (0.034)	P = 0.0774	0.94 (0.067)
Thoracic surgery	0.872 (0.391)	1.151 (0.470)	1.13 (0.654)	P = 0.0903	0.987 (0.751)
Trauma	0.997 (0.987)	1.057 (0.761)	1.265 (0.222)	P = 0.4373	1.02 (0.562)
Urology	0.827 (0.117)	1.105 (0.462)	1.24 (0.215)	P = 0.0334	0.977 (0.339)
Vascular surgery	1.103 (0.488)	1.052 (0.788)	0.966 (0.857)	P = 0.8116	0.946 (0.030)

Note: RR = Risk ratio

doi:10.1371/journal.pone.0131093.t005

We show that from 2011–2013 (Fig 2) the number of active wells continues to rise exponentially. Although we do not have health care utilization data for 2012–2013, if our findings persisted into 2012–2013, it is possible that the association between cardiology inpatient prevalence rates and wells could only become stronger as a result of the increased number of wells (relative to 2007–2011).

The precise cause for the increase in inpatient prevalence rates within specific medical categories remains unknown. Given that our modeling approach cannot account for within zip code demographic changes over the study period, it is possible that some increases were due to an increased influx of subjects to a zip code. Since the inpatient prevalence rates were determined for subjects *who resided* within a zip code, transient UGOD workers whose address was not local were excluded. Thus, our data potentially may underestimate hospital use that excluded those who were not Pennsylvania residents. Further, our data were partitioned into active wells but it is impossible to associate a specific toxicant exposure to an increase in a specific disease category requiring hospitalization. Intriguingly, our findings partially support those of other studies performed in Colorado. Colburn et al. observed that more than 75% of the chemicals used during natural gas operations may affect skin and respiratory systems, as well as other organs [23]. Another study in Colorado also supports our findings in neonatology. McKenzie et al. estimate that being within 10 miles of a gas well significantly increased the odds of having a congenital heart defect by 1.3 as well as the odds of having neural tube defects by two-fold, compared to not being within 10 miles of a gas well [12]. A recent study by Lanki et al. determined that living close to busy traffic was associated with increased C-reactive protein (CRP) concentrations, which is a known risk factor for cardiovascular diseases [24]. This supports our results for cardiology, given the increased truck traffic that comes with increased hydro-fracking activity.

Despite our findings that hospitalization use and active well number are directly associated within specific medical categories, there are limitations to our study. Our study examined a relatively short time interval. Whether our findings will be validated over longer periods of observation remains unclear. To have any association within a brief time frame may forebode greater negative health effects over time. Furthermore, with our limited time frame and data, the functional relationship for the association between well density and inpatient prevalence rates was heavily dependent on many extreme values, which make up less than 1% of the total observations. This motivated the quantile analysis. However, there are clear disadvantages to this approach. By partitioning a continuous variable, we inherently lose information. Furthermore, while we can make inference on moving among quantile levels, we cannot make inference for specific increases in well density. The quantile levels were also somewhat arbitrary, characterized as no wells/km^2, a "low" amount of wells/km^2, a "medium amount of wells/km^2, and a "high" amount of wells/km^2. Another possible limitation is that our analyses only considered a zip code "exposed" to wells if there were wells within that specific zip code. A zip code with no wells, however, could neighbor another zip code that has many wells. Accordingly, the association between wells and inpatient prevalence rates may be underestimated. Future work will incorporate a spatial aspect, such that the proximity to exposure (wells) is better addressed. Another limitation is that this study, given that we use hospital discharge data, does not include any information on morbidity or mortality. However, a future study that assesses the association between morbidity/mortality and wells would be interesting to explore.

Despite these limitations, our findings may have a significant impact on the consequences of UGOD on health care delivery and policy. For the number of wells analyses, it is useful to consider specific increases in wells, given that the risk ratio associated with the number of wells predictor is in terms of a one unit increase in number of wells. Specifically, consider an increase of 25 wells, which is the observed mean number of wells from our data. For example, if some zip code had an additional 25 wells, we would expect cardiology inpatient prevalence rates to increase by 2% for that zip code. Considering the quantile analyses, if a zip code went from having zero wells to having greater than 0.79 wells/km^2 (79 wells for each 100 km^2), we would expect cardiology inpatient prevalence rates to increase by 27% for that zip code. If a zip code went from having no wells to having between 0.17 to 0.79 wells/km^2, we then would expect a 14% increase in cardiology inpatient prevalence rates for that zip code. Notably, 18 zip codes had greater than 0.79 wells/km^2, primarily in 2010 and 2011, indicating that each of these zip codes could have had an excess of 27% in cardiology inpatient prevalence rates for each year they had greater than 0.79 wells/km^2. Furthermore, while dermatology and neonatology were not strictly significant after using a Bonferroni correction, there is evidence that dermatology and neonatology inpatient prevalence rates were also positively associated with wells. From the number of wells analyses, if a zip code had an additional 25 wells, we would expect dermatology and neonatology inpatient prevalence rates to increase by 3% and 4%, respectively. Similarly, from the quantile analyses, if a zip code went from having no wells to having greater than 0.79 wells/km^2, we would expect dermatology inpatient prevalence rates to increase by 45% for that zip code.

For most medical categories and overall, given the non-significant year risk ratios from Tables 4 and 5, inpatient prevalence rates remained relatively stable between 2007 and 2011. However, within the medical categories of gynecology and orthopedics, inpatient prevalence rates are expected to decrease each year by around 13–14% and 3–4%, respectively. Despite this surprising result, it is unclear why gynecology and orthopedics inpatient prevalence rates are decreasing each year. It is unlikely that these decreasing rates are related to the increased hydro-fracking activity.

To put into the context the potential burden of hydro-fracking on cardiology hospitalizations, consider the zip codes which exceeded 0.79 wells/km^2 (Q3wells). In total, from 2007 to 2011, three zip codes had $>$0.79 wells/km^2 in 2009, 10 zip codes had $>$0.79 wells/km^2 in 2010, and 18 zip codes had $>$0.79 wells/km^2 in 2011. Some zip codes had $>$0.79 wells/km^2 in multiple years, and in total, there were 18 unique zip codes that achieved $>$0.79 wells/km^2 at least once. Of these 31 year/zip code observations, the mean cardiology inpatient prevalence rate was 2.17, the mean number of cardiology inpatient visits was 44.74, and the mean population was 2190. Given the model results from Table 5, if these same observations had no wells, we would have expected the mean cardiology inpatient prevalence rate to be 2.17/1.27 = 1.71. Thus, the expected mean number of cardiology inpatient visits, assuming the mean population, would be 1.71*2190/100 = 37.46. However, this is a slight simplification, since each zip code has a different population. We omit the zip code specific populations to preserve zip code anonymity, but when using zip code specific populations, the expected mean number of cardiology inpatient visits, if these zip codes had no wells, would be 35.23. This means that on average, for any year that a zip code exceeded 0.79 wells/km^2, we would expect an excess of 44.74–35.23 = 9.51 cardiology inpatient visits, compared to if there were no wells. Note that this excess is for a single zip code for a single year in which the zip code exceeded 0.79 wells/km^2 (this occurred 31 times). A similar exercise shows that for zip codes in the Q2wells range (36 observations total), we would expect on average an excess of 8.13 cardiology inpatient rates. This again is for a single zip code for a single year in which the zip code had $>$0.168 wells/km^2 but $\leq$0.79 wells/km^2. However, from the model results in Table 5, zip codes with $>$1 well are in general expected to have increased cardiology inpatient prevalence rates, relative to having no wells. With an inpatient stay costing on average $30K, this poses a significant economic health burden to the Commonwealth of PA.

In summary, hydraulic fracturing as determined by well number or density had a significant association with cardiology inpatient prevalence rates, while well density had a significant association with neurology inpatient prevalence rates. While the clinical significance of the association remains to be shown, UGOD has just begun in Pennsylvania, and thus observing a significant association over this short time is remarkable. Further studies are warranted to compare toxicant exposure to number of wells and inpatient and outpatient studies. Our study also supports the concept that health care utilization should be factored into the value (costs and benefits) of hydraulic fracturing over time.

Supporting Information

S1 Table. ICD-9 diagnosis codes and MSDRGs used in this study. These data are partitioned into three tabs: ICD-9 diagnosis codes, MSDRGs and MSDRG product lines included.
(XLSX)

Author Contributions

Conceived and designed the experiments: TJ GLG MH PS TMP KJP RAP. Performed the experiments: TJ GLG MN SC BY MS MH PS NF TMP JR KJP RAP. Analyzed the data: TJ GLG MN SC BY MS NF TMP JR KJP RAP. Contributed reagents/materials/analysis tools: TJ GLG MN SC BY MS MH PS NF TMP JR KJP RAP. Wrote the paper: TJ GLG MN SC BY MS MH PS NF TMP JR KJP RAP.

References

1. IER (2012) U.S. Leads World in Natural Gas Production from Hydraulic Fracturing. Available: http://instituteforenergyresearch.org/analysis/u-s-leads-world-in-natural-gas-production-from-hydraulic-fracturing/#_edn1.
2. EIA (2014) EIA. Annual Energy Outlook 2014. Available: http://www.eia.gov/forecasts/aeo/mt_naturalgas.cfm.
3. Lampe DJ, Stolz JF. Current perspectives on unconventional shale gas extxraction in the Appalachian basin. J Environ Sci Health. 2015; 50: 434–446.
4. Schnoor JL. Shale gas and hydrofracking. Environ Sci Technol. 2012; 46: 4686. doi: 10.1021/es3011767 PMID: 22480196
5. Earthworks. Earthworks Hydraulic Fracturing 101. Available: http://www.earthworksaction.org/issues/detail/hydraulic_fracturing_101#.VSP49PnF98E.
6. EPA (2014) The Process of Hydraulic Fracturing. Available: http://www2.epa.gov/hydraulicfracturing/process-hydraulic-fracturing.
7. Jackson RB, Vengosha A, Darraha TH, Warnera NR. Increased stray gas abundance in a subset of drinking water wells near Marcellus shale gas extraction. Proc Natl Acad Sci USA. 2013; 110: 11250–11255. doi: 10.1073/pnas.1221635110 PMID: 23798404
8. Gross SA, Avens HJ, Banducci AM, Sahmel J, Panko JM, Tvermoes BE. Analysis of BTEX groundwater concenrations from surface spills associated with hydraulic fracturing operations. J Air Waste Manage Assoc. 2013; 63: 424–432.
9. Warner NR, Christie CA, Jackson RB, Vengosh A. Impacts of shale gas wastewater disposal on water quality in western Pennsylvania. Environ Sci Technol. 2013; 47: 11849–11857. doi: 10.1021/es402165b PMID: 24087919
10. Adgate JL, Goldstein BD, McKenzie LM. Potential public health hazards, exposures and health effects from unconventional natural gas development. Environ Sci Technol. 2014; 48: 8307–8320. doi: 10.1021/es404621d PMID: 24564405
11. Macy PG, Breech R, Chernaik M, Cox C, Larson D, Thomas D, et al. Air concentrations of volatile compounds near oil and gas production: a community-based exploratory study. Environmental Health. 2014; 13: 82. doi: 10.1186/1476-069X-13-82 PMID: 25355625
12. McKenzie LM, Guo R, Witter RZ, Savitz DA, Newman LS, Adgate JL. Birth outcomes and maternal residential proximity to natural gas development in rural Colorado. Environ Health Perspect. 2014; 122: 412–417. doi: 10.1289/ehp.1306722 PMID: 24474681
13. McKenzie LM, Witter RZ, Newman LS, Adgate JL. Human health risk assessment of air emissions from development of unconventional natural gas resources. Sci Total Environ. 2012; 424: 79–87. doi: 10.1016/j.scitotenv.2012.02.018 PMID: 22444058
14. Penning TM, Breysse PN, Gray K, Howarth M, Yan B. Environmental Health Research Recommendations from the Inter-Environmental Health Sciences Core Center Working Group on Unconventional Natural Gas Drilling Operations. Environ Health Perspect. 2014; 122: 1155–1159. doi: 10.1289/ehp.1408207 PMID: 25036093
15. Rabinowitz PM, Slizovskiy IB, Lamers V, Trufan SJ, Holford TR, Dziura JD, et al. Proximity to natural gas wells and reported health status: Results of a household survey in Washington County, Pennsylvania. Environ Health Perspect. 2014: Advance Publication 10 September 2014, http://dx.doi.org/2010.1289/ehp.1307732.
16. PA Department of Environmental Protection (2013). Marcellus Shale Development. Available: http://www.elibrary.dep.state.pa.us/dsweb/Get/Document-97683/0100-FS-DEP4217.pdf
17. PA Department of Environmental Protection (2013). 2013 Oil and Gas Annual Report. Available: http://www.portal.state.pa.us/portal/server.pt/community/annual_report/21786
18. MarcellusGas.Org www.marcellusgas.org—Information related to Pennsylvania deep gas well activity.
19. US Government (2010). 2010 Census Data. Available: http://www.census.gov/2010census/data/.

20. PA Departent of Environmental Protection (2014). PA DEP Oil & Gas Reporting Website. Available: https://www.paoilandgasreporting.state.pa.us/publicreports/Modules/DataExports/DataExports.aspx. Pennsylvania Department of Environmental Protection.

21. Wooldridge JM. Distribution-free estimation of some nonlinear panel data models. Journal of Econometrics. 1999; 20: 77–97.

22. Bernstein AB, Hing E, Moss AJ, Allen KF, Siller AB, Tiggle RB. (2003) Health Care in America: Trends in Utilization. In: U.S. Department of Health and Human Services CfDCaP, National Center for Health Statistics, editor. Hyattsville, Maryland: National Center for Health Statistics.

23. Colburn T, Kwiatkowski C, Schultz K, Bachran M. Natural gas operations from a public health perspective. Hum Ecol Risk Assess. 2011; 17: 1039–1056.

24. Lanki T, Hampel R, Tittanen P, Andrich S, Beelen R, Brunekreef B, et al. Air pollution from road traffic and systemic inflammation in adults: a cross-sectional analysis in the European ESCAPE Project. Environ Health Perspect. 2015, Advance Publication 27 March 2015.

ATTACHMENT C

STUDY 16

The Extractive Industries and Society 1 (2014) 124–126

Contents lists available at ScienceDirect

The Extractive Industries and Society

journal homepage: www.elsevier.com/locate/exis

Viewpoint

Distance: A critical aspect for environmental impact assessment of hydraulic fracking

Qingmin Meng [a,*], Steve Ashby [b]

[a] Department of Geosciences, Mississippi State University, MS 39762, USA
[b] Northern Gulf Institute, Mississippi State University, MS 39762, USA

ARTICLE INFO

Article history:
Received 5 June 2014
Received in revised form 11 July 2014
Available online 2 August 2014

Keywords:
Hydraulic fracking
Environmental risks
Spatial analysis
Distance
Methane migration

ABSTRACT

Public concerns about hydraulic fracking are growing and scientists continue to analyze and evaluate its associated environmental impacts. However, a rigorous spatial analysis of environmental impacts is necessary to provide a perspective on risk based on proximity to fracking wells. This comment describes the environmental impacts of fracking within a spatial context. It emphasizes five key points: (1) the closer to a hydraulic fracking well, the higher the risk of groundwater and drinking water well contamination; (2) residents living nearest to a fracking well experience a higher human health risk due to exposure to gas emissions during the fracking process; (3) huge and high density gas emissions are detected and recorded close to fracking wells; (4) fracking induces seismicity and small earthquakes are recorded close to fracking wells; and (5) hydraulic fracking directly changes local environment and landscape characteristics. Spatial impact assessments are critical for improving understanding of the impacts of hydraulic fracking on the environment and society.

1. Introduction

Hydraulic fracking, also called fracking, is the process of extracting natural gas from shale rock layers within the earth. Specifically, horizontal drilling combined with traditional vertical drilling allows injection of highly pressurized fracking fluids into the shale layers to create new channels within the rock, from which natural gas is released at much higher rates than traditional drilling. Hydraulic fracking yields more than one-half of US natural gas supply and is transforming energy supplies in the United States (Jackson et al., 2013). For example, in January of 2013, the daily production of methane in the United States was 2×10^9 m^3, more than a 30% increase from 2005 (USEIA, 2013). Fracking gas production in Northeastern Pennsylvania now exceeds 2 billion cubic feet per day, up from 0.4 billion cubic feet per day in early 2010. In Southwestern Pennsylvania, it is close to 1 billion cubic feet per day, more than three times the production of early 2010 (WhatIsFracking, 2014).

Environmental concerns about hydraulic fracking are growing (Osborn et al., 2011; Schmidt 2011). These concerns include changes in air quality (Petron et al., 2012), human health risks for populations living near fracking wells (Schmidt, 2011), and the potential persistence of pollutants in groundwater and drinking water in close proximity to hydraulic fracking sites. For example, hydraulic fracking from the Marcellus Shale in the Appalachian Basin of the Northeastern United States has raised concerns about potential environmental pollution (Kerr, 2010; Kargbo et al., 2010). Methane migration to groundwater, drinking water wells, and the atmosphere (Howarth et al., 2011a; Osborn et al., 2011; Jiang et al., 2011) is of particular concern. Additional concerns include induced seismicity associated with fluid injection into deep wells (Ellsworth et al., 2012), epicenters of small earthquakes within an approximate 1 km radius to the fracking well (Kim, 2013), and surface environmental and landscape changes (Meng, 2014).

Recent studies have failed to feature any rigorous spatial analysis but have suggested that spatial dimensions of environmental impacts exist, and are largely a function of distance to fracking sites. It is therefore time for decision makers and scientists to pay closer attention to the spatial planning of hydraulic fracking, prioritizing the issue of distance to a hydraulic fracking well in environmental impact assessments. This is imperative, given the rapid rise in number of sites and their close proximity to water supplies and communities.

* Corresponding author. Tel.: +1 662 325 3915.
E-mail address: qmeng@geosci.msstate.edu (Q. Meng).

http://dx.doi.org/10.1016/j.exis.2014.07.004
2214-790X/

2. Distance is all-important

Residents living within 0.8 km from a gas well are at higher risks of health effects than residents living beyond this distance (Mckenzie et al., 2012). Mckenzie et al. (2012) and Coons and Walker (2008) found that significant gas emissions exist close to a gas well (<0.8 km). Methane concentrations in drinking water wells within 1 km of a gas fracking well can reach potential explosion levels (Osborn et al., 2011). Methane concentrations are six times higher and ethane concentrations were found to be 23 times higher at residences within 1 km of a shale gas fracking site compared with concentrations at distant residences, and additionally, propane was detected in water wells within approximately 1 km of a fracking well site (Jackson et al., 2013). Subsurface and surface pathways exist although specific pathways of methane migration are not easily identifiable. Traces of ethane (C_2H_6) with microbial methane (CH_4) and a range of C and H isotopic compositions of CH_4 indicated that sub-surface pathways exist and gas mixtures are found in groundwater (Revesz et al., 2010).

Vidic et al. (2013) carried out an important review of the effects of shale gas development on regional water quality. However, reviews of methane migration are limited and tell very little. For example, Vidic et al. (2013) reported findings from a study of 48 water wells for pre- and post-drilling water chemistry that showed no statistical differences in dissolved methane before or shortly after drilling, and distance to drilling sites was not found to be significant. However, the authors did not take into account that among the 48 water wells, at 16 of the sites, only drilling—and no fracking—had occurred. Furthermore, 28% of the 33 water supply owners who reported changes to their water supply after drilling were located within 3,000 feet (0.914 km) of a Marcellus gas well (Boyer et al., 2011).

Based on a series of studies conducted by the EPA and other scientists, Howarth et al. (2011b) concluded that 3.6–7.9% of lifetime shale gas production migrates to the atmosphere through venting or leaking over a well is lifetime and that 1.9% of the total gas production is emitted as methane through well completion. For example, methane emitted during flow-back was determined to be 6800×10^3 M^3 with a per day rate of 680×10^3 M^3 for a fracking well in Louisiana, and calculated to be 370×10^3 M^3 with a per day rate of 41×10^3 M^3 for a shale gas well in Texas. Caulton et al. (2014) identified and quantified large emissions with an average of 34 g CH_4/s (2.937 ton/day) per well from seven hydraulic fracking pads in the drilling phase. These emissions are 2–3 orders of magnitude greater than the estimates formulated by the US Environmental Protection Agency. This methane can migrate to soils and open water through both wet and dry deposition. More data and studies are needed, however, to identify specific pathways for methane migration and how it impacts local air and water quality.

Construction of hydraulic fracking wells alters the local environment and land surface. Land clearing, excavating and grading, pad construction, pipeline and utility installation, related road construction, sump hole excavation, and hydroseeding as well as soil stabilization are the main construction activities that impact the local landscape. These activities also result in a much larger area being impacted than a conventional gas or oil drilling well. Additionally, the excavation of natural gas and oil resources from shale typically requires much more water. In the fracking process, fluids are forced under high pressure into the well, and the shale surrounding the borehole is fractured in order to liberate more gas from the low permeability shale gas reservoirs.

Meng (2014) has modeled the impacts of hydraulic fracking based on environmental and landscape variables. Statistical diagnostics of spatial logistic regression models show that elevation, slope, and land cover are significant environmental and landscape variables. A location with steeper slopes is less likely to become a fracking site. Sites at higher elevations are more likely to be occupied by fracking wells.

3. Conclusion

Hydraulic fracking has the potential to cause significant impact to local environments and landscapes. The closer a site is to a hydraulic fracking well, the greater the hydraulic impacts associated activity will have on the surrounding environment. There is a higher probability of the groundwater and drinking water wells which are located within 1 km of a fracking having been polluted by gas and fracking chemicals. The risk to human health is especially high among populations located within 0.8 km of a fracking well. High density gas emissions typically persist in the surface air close to fracking wells, and small earthquakes have been detected close to a deep fluid injection well. It is time to pay attention to concerns in order to develop a more comprehensive understanding and assessment of the environmental impacts of hydraulic fracking.

Acknowledgement

The authors would like to thank Professor Gavin Hilson for his comments and editing, which enhanced the quality of this viewpoint.

References

Boyer, E.W., Swistock, B.R., Clark, J., Madden, M., Rizzo, D.E., 2011.In: The impact of Marcellus Gas Drilling on Rural Drinking Water Supplie. The Center for Rural Pennsylvania, Pennsylvania General Assemble. www.rural.palegislature.us/documents/reports/Marcellus_and_drinking_water_2011_rev.pdf.

Caulton, D.R., Shepson, P.B., Santoro, R.L., Sparks, J.P., Howarth, R.W., Ingraffea, A.R., Cambaliza, M.O.L., Sweeney, C., Karion, A., Davis, K.J., Stirm, B.H., Montzka, S.A., Miller, B.R., 2014. Toward a better understanding and quantification of methane emissions form shale gas development. PNAS 111 (17), 6237–6242.

Coons, T., Walker, R., 2008.In: Community Health Risk Analysis of Oil and Gas Industry Impacts in Garfield County. . http://www.garfield-county.com/public-health/documents/1._COMMUNITY_HEALTH_RISK_ANALYSIS-(Complete_Report_16MB).pdf.

Ellsworth, W.L., Hickman, S.H., Llesons, A.L., Mcgarr, A., Michael, A.J., Rubinstein, J.L., 2012. Are Seismicity Rate Changes in the Midcontinent Natural or Manmade? US Geological Survey, Menlo Park, CA.

Howarth, R.W., Ingraffea, A., Engelder, T., 2011a. Natural gas: should fracking stop? Nature 477, 271–275.

Howarth, R.W., Santoro, R., Ingraffea, A., 2011b. Methane and the greenhouse-gas footprint of natural gas from shale formations. Climatic Change 106, 679–690.

Jackson, R.B., Vengosh, A., Darrah, T.H., Warner, N.R., Down, A., Poreda, R.J., Osborn, S.G., Zhao, K., Karr, J.D., 2013. Increased stray gas abundance in a subset of drinking water wells near Marcellus shale gas extraction. Proc. Natl. Acad. Sci. U.S.A. 110, 11250.

Jiang, M., Griffin, W.M., Hendrickson, C., Jaramillo, P., VanBriesen, J., Venkatesh, A., 2011. Life cycle greenhouse gas emissions of Marcellus shale gas. Environ. Res. Lett. 6, 034041.

Kerr, R.A., 2010. Natural gas from shale bursts onto the scene. Science 328, 1624–1626.

Kargbo, D.M., Wilhelm, R.G., Campbell, D.J., 2010. Natural gas plays in the Marcellus Shale: challenges and potential opportunities. Environ. Sci. Technol. 44, 5679–5684.

Kim, W.Y., 2013. Induced seismicity associated with fluid injection into a deep well in Youngstown, Ohio. J. Geophys. Res. Solid Earth 118, 1L 13.

Mckenzie, L.M., Witter, R.Z., Newman, L.S., Adgate, J.L., 2012. Human health risk assessment of air emissions from development of unconventional natural gas resources. Sci. Total Environ. 424, 79–87.

Meng, Q., 2014. Modeling and prediction of natural gas fracking pad landscapes in the Marcellus Shale region, USA. Landscape Urban Plann. 121, 109–116.

Osborn, S.G., Vengosh, A., Warner, N.R., Jackson, R.B., 2011. Methane contamination of drinking water accompanying gas-well drilling and hydraulic fracturing. Proc. Natl. Acad. Sci. U.S.A. 108, 8172.

Petron, G., Frost, G., Miller, B.R., Hirsch, A.I., Montzka, S.A., Karion, A., Trainer, M., Sweeney, C., Andrews, A.E., Miller, L., Kofler, J., Bar-Ilan, A., Dlugokencky, E.J., Patrick, L., T. Moore Jr., C.T., Ryerson, T.B., Siso, C., Kolodzey, W., Lang, P.M., Conway, T., Novelli, P., Masarie, K., Hall, B., Guenther, D., Kitzis, D., Miller, J., Welsh, D., Wolfe, D., Neff, W., Tans, P., 2012. Hydrocarbon emissions characterization in the Colorado Front Range: a pilot study. J. Geophys. Res. 117 (D4), D04304.

Revesz, K.M., Breen, K.J., Baldassare, A.J., Burruss, R.C., 2010. Carbon and hydrogen isotopic evidence for the origin of combustible gases in water-supply wells in north-central Pennsylvania. Appl. Geochem. 25, 1845–1859.

Schmidt, C.W., 2011. Blind rush? Shale gas boom proceeds amid human health questions. Environ. Health Perspect. 119 (8), A348–A353.

USEIA, 2013. Natural Gas Monthly March 2013. US Energy Information Administration, Washington, DC DOE/EIA 0130 (2013/03).

Vidic, R.D., Brantley, S.L., Vandenbossche, J.M., Yoxtheimer, D., Abad, J.D., 2013. Impact of Shale Gas Development on Regional Water Quality. Science 340, 1235009, http://dx.doi.org/10.1126/science.1235009.

WhatIsFracking, 2014. Economic Opportunity, www.what-is-fracking.com/economic-opportunity/ (accessed on 12.06.14).

ATTACHMENT C

STUDY 17

frontiers | Frontiers in Epidemiology

TYPE Original Research
PUBLISHED 19 June 2024
DOI 10.3389/fepid.2024.1379271

Check for updates

OPEN ACCESS

EDITED BY
Alexandra Schneider,
Helmholtz Association of German Research Centres (HZ), Germany

REVIEWED BY
Matteo Renzi,
Local Health Authority Rome 1, Italy
Massimo Stafoggia,
Regional Health Service of Lazio, Italy

*CORRESPONDENCE
Lisa M. McKenzie
✉ lisa.mckenzie@cuanschutz.edu

RECEIVED 30 January 2024
ACCEPTED 03 June 2024
PUBLISHED 19 June 2024

CITATION
McKenzie LM, Allshouse WB, Abrahams B and Tompkins C (2024) Oil and gas development exposure and atrial fibrillation exacerbation: a retrospective study of atrial fibrillation exacerbation using Colorado's all payer claims dataset.
Front. Epidemiol. 4:1379271.
doi: 10.3389/fepid.2024.1379271

Oil and gas development exposure and atrial fibrillation exacerbation: a retrospective study of atrial fibrillation exacerbation using Colorado's all payer claims dataset

Lisa M. McKenzie[1]*, William B. Allshouse[1], Barbara Abrahams[2] and Christine Tompkins[3]

[1]Department of Environmental and Occupational Health, Colorado School of Public Health, University of Colorado Anschutz Campus, Aurora, CO, United States, [2]Department of Cardiology, University of Colorado School of Medicine, University of Colorado Anschutz Campus, Aurora, CO, United States, [3]Division of Electrophysiology, Emory University, Atlanta, GA, United States

Introduction: Emerging risk factors for atrial fibrillation (AF) incidence and episodes (exacerbation), the most common and clinically significant cardiac arrhythmia, include air and noise pollution, both of which are emitted during oil and natural gas (O&G) well site development.
Methods: We evaluated AF exacerbation risk and proximity to O&G well site development by employing a novel data source and interrupted time-series design. We retrospectively followed 1,197 AF patients living within 1-mile of an O&G well site (at-risk of exposure) and 9,764 patients living >2 miles from any O&G well site (unexposed) for AF claims in Colorado's All Payer Claims Dataset before, during, and after O&G well site development. We calculated AF exacerbation risk with multi-failure survival analysis.
Results: The analysis of the total study population does not provide strong evidence of an association between AF exacerbation and proximity to O&G wells sites during (HR = 1.07, 95% CI: 0.94, 1.22) or after (HR = 1.01, 95% CI: 0.88, 1.16) development. However, AF exacerbation risk differed by patient age and sex. In patients >80 years living within 0.39 miles (2,059 feet) of O&G well site development, AF exacerbation risk increased by 83% (HR = 1.83, 95% CI: 1.25, 2.66) and emergency room visits for an AF event doubled (HR = 2.55, 95% CI: 1.50, 4.36) during development, with risk increasing with proximity. In female patients living within 0.39 miles of O&G well site development, AF exacerbation risk increased by 56% percent (95% CI: 1.13, 2.15) during development. AF exacerbation risk did not persist past the well development period. We did not observe increased AF exacerbation risk in younger or male patients.
Discussion: The prospect that proximity to O&G well site development, a significant noise and air pollution source, may increase AF exacerbation risk in older and female AF patients requires attention. These findings support appropriate patient education to help mitigate risk and development of mitigation strategies and regulations to protect the health of populations in O&G development regions.

KEYWORDS

atrial fibrillation, oil and natural gas development, cohort study, environmental epidemiology, hydraulic fracturing, air pollution, noise pollution

Introduction

Atrial fibrillation (AF), the most common and clinically significant cardiac arrhythmia, impairs quality of life and substantially elevates stroke, systemic thromboembolism and heart failure risk (1, 2). The incidence and prevalence of AF are increasing (1–5). Adults aged >40 years of age have a 25% lifetime risk of developing AF (1). There are 9.3 million American's living with this chronic, dangerous, and costly condition contributing to an estimated 130,000 deaths and $6 billion in health care costs per year (6).

While knowledge on AF etiology is sparse, there are several known AF risk factors, including biological sex, advancing age, and co-morbidities (1), as well as emerging environmental risk factors including air and noise pollution (7–11). Several epidemiological studies have indicated that the risk of AF incidence increases with increasing levels of air pollutants, including particulate matter ≤2.5 micrometers, ($PM_{2.5}$), nitrogen oxides (NO_x), and ozone (8, 12–18), as well as higher exposure to traffic and railway noise (9, 19, 20). Additionally, studies have observed the risk of AF episodes (exacerbation) increases with increasing levels of air pollutants and noise (13, 14, 21, 22). In general, adverse cardiovascular effects are observed when audible noise levels exceed 50 A-weighted decibels (dBA) (23). Studies also suggest that nocturnal noise, which disrupts the normal sleep cycle, may be associated with greater health consequences than daytime noise (24–26). Clinically, chronic sleep deprivation is associated hypertension (27) and cardiovascular disease (28) which are firmly established and modifiable risk factors for AF (1).

One significant source of both air and noise pollution is the development of oil and natural gas (O&G) well sites. Between 2011 and 2014, 25,000–35,000 O&G well sites were developed annually in the United States (US) exceeding 150,000 total new well sites as of 2019 (29). This resulted in an extensive dispersion of O&G well sites across populated areas, with over 17 million people living within one mile of an O&G well (30). In Colorado, more than 378,000 people live within 1-mile of an O&G well site, with the densest development northeast of Denver (31). Air and noise pollution emitted during development of O&G well sites potentially impact all individuals residing near the sites (32).

As described elsewhere, modern O&G well site development is a complex, industrial process (33). Diesel-powered equipment, trucks, and generators continuously emit air pollutants and noise; on-site storage tanks, valves and pipes also emit air pollutants (34–36). Audible noise levels of 69 dBA and low frequency noise of 80 C-weighted decibels (dBC) have been reported during O&G well site development (35, 37). During development of 22-well O&G site in Colorado, 1–16 diesel trucks per hour travelled to and from the site, concentrations of $PM_{2.5}$ more than doubled, and noise measurements exceeded 50 dBA day and night, within1,288 feet of the site (36).

It is not known if noise and air pollution emitted from O&G well site development exacerbates AF in the large and growing population living near these sites. We are not aware of any studies on this topic. However, studies indicate that living near O&G well sites may impact cardiac conditions associated with AF. Proximity to O&G well sites may affect *in-utero* heart development (38–42), increase hospitalizations for heart failure (43) in acute myocardial infarction patients (44), and increase augmentation index and blood pressure (45). Our objective is to determine if the burden of AF increases in AF patients living near O&G well site development and identify susceptible subpopulations by employing a novel time-series design and data source in a large population of AF patients using specific O&G metrics. Because air pollution and noise emissions persist in the production period following well site development, we also determine AF exacerbation increases (or persists) after the well site is developed.

Methods

We retrospectively followed 10,961 AF patients in Colorado's All Payer Claims Dataset (COAPCD) before, during, and after development of O&G well sites. Using both an interrupted time series (ITS) and controlled interrupted time series (CTIS) design (Figure 1) (46, 47), we evaluated if living near O&G well site during development exacerbates AF and if AF exacerbation persists after development of the site. We selected an ITS design because of the limited co-variate information available in Colorado's All Payer Claims Dataset (APCD). Because ITS is based on observation of a single population over time, it accounts for between group differences, such as unmeasured confounding, as well as within group characteristics that change slowly over time, secular changes, random fluctuations from one point to the next and regression to mean (46). To control for time-varying trends which do not form part of the underlying trend (e.g., seasonal, regional scale environmental events, and natural progression of AF), we also performed a CITS by adding an unexposed group as recommended by Bernal et al. (47) Per these recommendations, we included and reported results from both the ITS and CITS to provide a greater degree of confidence that an observed association between proximity to development of an O&G well site and AF exacerbation is causal (47). For example, if the CITS analysis indicates an association, but the ITS does not, then there may have been an event affecting AF in the control population that did not affect the population living within one mile. The Colorado Multiple Institutional Review Board approved our study (IRB Protocol Number 17–0692).

Study population

We selected our cohort from the COAPCD, administered by The Center for Improving Value in Health Care. The COAPCD represents approximately 65% of Colorado's fully insured population including claims data from commercial health plans (large group, small group, and individual), Medicare, and Colorado's Medicaid Program beginning in 2009. We included patients in the COAPCD aged 18–100 years with a complete street address that we could geocode, living in a Colorado county with at least one O&G well site developed between 2010 and

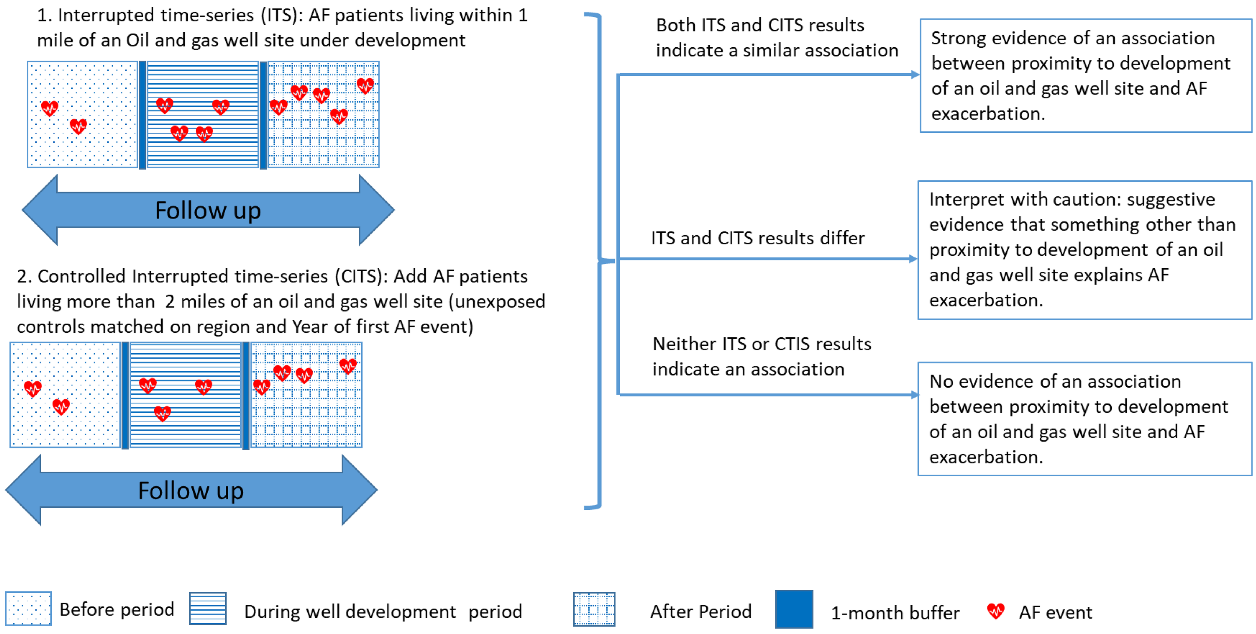


FIGURE 1
Interrupted time-series study design and assignment of before, during, and after oil and gas well site development period. AF, atrial fibrillation [adapted from Bernal et al. (47)].

2017, and at least one principal diagnosis code for AF or atrial flutter (AFl) between January 1, 2009 and December 31, 2017. From this population, we *a priori* selected patients at- risk of exposure to air and noise pollution emitted during development of an O&G site (herein referred to as exposed patients) and an unexposed population as follows.

We calculated the distance between each patient's geocoded address and the nearest O&G well site developed between 2010 and 2017 using ArcGIS Desktop 10 as described in the exposure section. We defined patients at-risk of exposure (here to in referred to as at-risk patients) as living within one mile of an O&G site based on documented noise and odor complaints, recent risk assessments, and monitoring studies indicating the potential for air and noise pollution associated with O&G well sites to impact people living within one mile (36, 48–50), as well as a robust literature supporting the use of proximity to O&G well sites as a proxy for exposure (51). Because weather, major air pollution events, and other temporal events that could exacerbate AF vary by region and AF severity may worsen over time, analysis of an location control population was necessary (47). The location control population (here to in referred to as unexposed patients) should be a population not a risk for exposure to air and noise pollution emitted from an O&G site. We selected our unexposed population from AF patients that had no O&G well sites within two miles of their home by frequency matching each at-risk patient to 13 unexposed patients by geographical region to control for regional temporal events (Supplementary Material Table S1) and year of first AF claim in the COAPCD to control for progression of AF severity. Because the spatial extent of stressors from O&G site development is not well understood and may extend beyond 1-mile, we excluded patients living 1–2 miles from an O&G site to clearly distinguish the possibility of exposure to O&G well site development stressors in at-risk patients from unexposed patients.

Exposure

We geocoded street addresses in ArcGIS Desktop 10 using Census TIGER Address Range files from 2019 to create an address locator. For patients that could not be geocoded with ArcGIS Desktop 10, we completed a second geocoding pass with the Google Geocoding API. We obtained geocoded O&G well site locations for all O&G wells developed between 2010 and 2017, the number of wells at each well site, and the dates those wells were developed (spud date, first production date) from the Colorado Oil and Gas Information System (52).

Assuming the street address in the COAPCD is also the residential address, we temporally aligned each matched control set (up to 13 patients) to the development of the O&G well site within one mile of their matched at-risk patient's street address. We defined before, during, and after development periods as follows (Figure 1). The during development period begins on the drilling date (the spud date) of the first well on the site and ends on the first production date of the last well on the site. We then added a one-month buffer to the beginning and end of the during development period to account for well site construction activities prior to drilling and higher potential activities at the beginning of production. The before and after development periods each are equal to the length of the during development

10.3389/fepid.2024.1379271

period. The before period ends at the beginning of the one-month buffer period preceding the development period. The after period begins at the end of the one-month buffer following the development period.

Exclusions

We excluded patients living 1–2 miles from an O&G site (Figure 2). We next excluded 2,353 patients if the date of their first AF claim occurred later than the end of the after period because there was no evidence that the patient had AF during the follow up period. To reduce errors from unknown losses to follow up, we also excluded 5,626 patients without a claim (of any type) preceding the before period and succeeding the after period as defined in the exposure section. Our final population of 10,961 AF patients includes 1,197 at-risk and 9,764 control patients. A blinded review of claims for 1% of randomly selected at-risk and control patients confirmed 91% (90% at-risk and 94% control) patients were correctly identified as having a primary diagnosis of AF. Insufficient information was available in the remaining 9% of these patients to confirm a primary AF diagnosis.

Outcomes

We followed each patient from the beginning of their specific before period through the end of the after period (here to within referred as follow up) for incidence of an AF episode. We defined an AF episode as any claim, inpatient, outpatient, and emergency room, with a principal diagnostic code for AF or AFI (ICD-9-CM 427.3, or 427.31–2; ICD-10-CM I48.0–4, I48.9, I48.91, or I48.92), excluding AF diagnostic codes associated with an internal normalized ratio procedure (CPT4 85610, 93792, or R79.1). We considered occurrence of multiple AF diagnostic codes in one day or over consecutive days as one event. We also separately evaluated AF episodes associated with an emergency

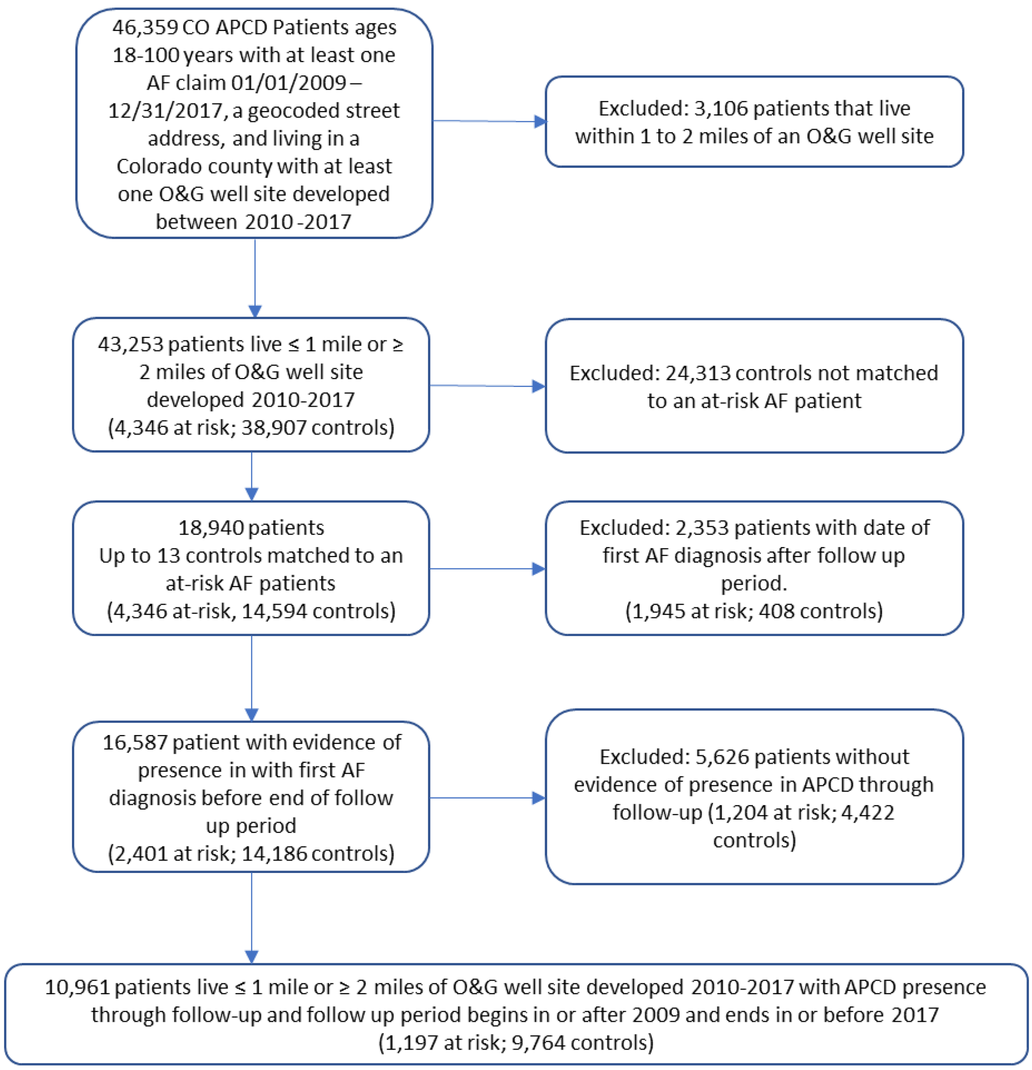


FIGURE 2
Selection of atrial fibrillation patients from Colorado all payers claim database.

room visit (Supplementary Material Table S2). AF episodes that occurred in the buffer months were not counted.

Statistical analyses

We tested the hypothesis that there is a larger increase in incidence of AF claims during or after development of an O&G site, compared to before development, for at-risk AF patients as follows. We analyzed AF exacerbation risk with a multi-failure survival analysis by applying a Cox proportional hazard model with a robust variance estimator and clustering at the individual patient level (53, 54), using an Efron method for ties (55). We retrospectively followed each patient through their specific follow up period. We first analyzed AF exacerbation risk for only the at-risk patients (ITS) (46, 47). We then analyzed AF exacerbation risk with both the at-risk and unexposed patients (CTIS) (47), by adding an interaction term between exposure (at-risk to control referent) and period (during and after, to before referent) to our model. Parallel trend analysis indicates no difference between the exposed and unexposed populations in the before period, indicating support for the parallel assumption in CITS analysis (Supplementary Material Table S3) (56). Strong evidence of an association between proximity to development of an O&G well site and AF exacerbation is indicated if the ITS and CITS analysis yield hazard ratios (HR) of similar size (Figure 1) (47). We adjusted our model for co-variates associated with AF (biological sex, age at first AF claim in COAPCD, elevation of residence, hypertension, and diabetes) (1), and exposure (duration of well development and geographical region). We considered the direction and magnitude of individual HRs and overarching trends, based on American Statistical Association guidance (57), in both analyses.

We then stratified our population by sex, age quartiles, presence of a co-morbidity (diabetes, hypertension) and geographical region to assess whether the results between groups (e.g., male vs. female) were systematically different. Additionally, we stratified our at-risk patients into distance quartiles to assess the effect of distance from the O&G site on AF exacerbation.

We performed several sensitivity analyses. We evaluated the effect of short and long periods of well development by excluding patients with well duration periods outside the 25th—75th percentile range (75–184 days). To evaluate the impact of potential change of residence over time, we performed an analysis on patients for whom we could confirm that the street address did not change through the follow up period. Because high altitude can exacerbate AF, we performed an analysis on patients living ≤6,000 feet above sea level.

Given the small sample sizes and exploratory nature of the stratified and sensitivity analyses, no adjustments were made for multiple comparisons. All analyses were carried out using SAS 9.4 (SAS Institute, Cary, NC).

Results

Our study population included 1,197 at-risk patients and 9,764 unexposed patients l (Table 1). At-risk patients were more likely to be male, diabetic, and hypertensive than unexposed patients. At-risk patients also had a longer duration in the COAPCD. However, the highest number of total AF claims was observed in the unexposed patients.

TABLE 1 Study population characteristics for All payer claims database patients aged 18–100 years living within one mile of an oil and gas well developed in Colorado between 2010 and 2017 or two or more miles from any Colorado oil and natural gas well site.

	At-risk: nearest well within one mile	Control: nearest well two or more miles
Total *N*	1,197	9,764
Age at first AF event in COAPCD (%)		
Greater than 80 years	27.6	28.2
74–80 years	23.5	23.8
66–73 years	30.1	27.2
<66 years	18.8	20.9
Sex (%)		
Male	53.0	50.7
Female	46.7	48.3
Missing	<1	<1
Diabetic (%)	34.9	30.5
Hypertensive (%)	83.1	77.8
Confirmed address over follow up (%)	73.2	72.2
Region		
East	89.9	89.5
Southwest	4.4	4.8
Northwest	5.7	5.6
Elevation of residence <6,500 feet	97.3	94.2
Emergency room visits (%)	30.2	30.1
Duration in COAPCD (days)		
Mean	3,023	2,983
Maximum	3,651	3,651
Minimum	605	317
Number of AF events		
Mean	14.2	13.8
Maximum	161	373
Minimum	0	0
Miles from O&G well site (*n*)		
0–0.39	299	–
>0.39–0.59	302	–
0.59–0.80	303	–
>0.8–1	293	–
Duration of O&G well site development (days)		
Mean	165	–
Maximum	844	–
Minimum	3	–

AF, atrial fibrillation; COAPCD, Colorado all payer claims dataset; O&G, oil and gas.

Table 2 presents the multi-failure survival analysis results for AF exacerbation. The analysis of the study population as a whole does not provide strong evidence of an association between AF exacerbation and proximity to O&G well site development. The ITS analysis indicates that AF exacerbation increases during (HR = 1.13, 95% CI: 0.99, 1.30) and after (HR = 1.19, 95% CI = 1.02, 1.39) well site development, compared to before well development in our total population of at-risk AF patients. With

TABLE 2 Results from multi-failure survival analysis for AF events: hazard ratios for an AF event during and after well development periods compared to before well development.

Analysis	Interrupted time series analysis[a]		Controlled interrupted time series	
	HR during[b] (95% CI)	HR after[b] (95% CI)	HR during[c] (95% CI)	HR after[c] (95% CI)
Total population (Main analysis)	1.13 (0.99, 1.30)	1.19 (1.02, 1.39)	1.07 (0.94, 1.22)	1.01 (0.88, 1.16)
>80 years	1.41 (1.09, 1.83)	1.19 (0.93, 1.52)	1.43 (1.13, 1.81)	0.99 (0.79, 1.23
74–80 years	1.09 (0.84, 1.42)	1.16 (0.84, 1.61)	1.11 (0.86, 1.44)	1.06 (0.79, 1.40)
66–73 years	1.08 (0.84, 1.40)	1.43 (1.02, 2.0)	0.88 (0.69, 1.12)	1.10 (0.82, 1.47
<66 years	0.83 (0.61, 1.17)	0.87 (0.65, 1.17)	0.84 (0.61, 1.17)	0.87 (0.65, 1.17)
Females	1.17 (0.97, 1.42)	1.14 (0.91, 1.44)	1.21 (1.00, 1.47)	1.02 (0.82, 1.27)
Males	1.09 (0.90, 1.31)	1.25 (1.02, 1.55)	0.94 (0.79, 1.11)	1.01 (0.86, 1.20)

AF, atrial fibrillation; CI, confidence interval; HR, hazard ratio.
[a]Does not include control patients.
[b]Adjusted for sex, age at first AF event, elevation of residence, duration of well development, hypertension, diabetes, and region, and exposure status.
[c]Adjusted for sex, age at first AF event, elevation of residence, duration of well development, hypertension, diabetes, and region, interaction between period (before, during, after development) and exposure status.

inclusion of unexposed patients in the CITS analyses this association attenuates towards the null during (HR = 1.07, 95% CI: 0.94, 1.22) and after (HR = 1.01, 95% CI: 0.88, 1.16) site development.

Stratified analyses indicates that age and possibly biological sex modify the risk of an AF event during O&G well site development (Table 2). In AF patients aged >80 years, ITS and CITS results for the during well development period are similar and indicate that risk for an AF event increases during well development, but not after development. In the during O&G development period, risk of an AF event increased by 43% (HR = 1.43, 95% CI: 1.13, 1.81) in at-risk patients aged >80 years. In younger patients, results attenuated towards the null. In female AF patients, ITS and CITS results for the during well development period are similar and indicate that risk for an AF event increases during well development, but not after development. The risk for an AF event in the during O&G development period increased by 21% (HR = 1.21, 95% CI: 1.00, 1.47) in at-risk female patients. No association was observed for male patients. Stratified analysis indicated that co-morbidities and region of residence did not modify risk (Supplementary Material Table S4).

In stratified analysis by distance quartile, ITS and CTIS results for the total population are similar and indicate increased risk of AF exacerbation during well site development in at-risk patients living within 0.39 miles (2,059 feet) and the increased risk does not persist after development of the well site. We observed a 35% increase in risk for AF events in at-risk patients living within 0.39 miles (2,059 feet) in the during well development period (HR = 1.35, 95% CI: 1.08, 1.69) (Figure 3, Supplementary Material Table S5). We did not observe associations at distances >0.39 miles for the total population. As in the main analysis, both age and sex modified the results. In patients aged >80 years living within 0.39 miles of an O&G development site, the risk of AF event increased by 83% during well development (HR = 1.83, 95% CI: 1.25, 2.66). Additionally, the results for patients aged >80 years indicate a trend of increasing risk of an AF event during well development as distance from the well site decreases and suggest the possibility of increased AF exacerbation risk up to 4,224 feet from the site. In female patients living within 0.39 miles of an oil and gas well development site, risk of AF event increased by 56% (95% CI: 1.13, 2.15) and 36% (95% CI: 0.89, 2.03) during and possibly after well development, respectively.

Table 3 presents the multi-failure survival analysis results for AF exacerbation with an emergency room visit. The analysis of the study population as a whole does not provide strong evidence of an association between AF exacerbation with an emergency room visit and proximity to O&G well site development. The ITS analysis indicates that AF exacerbation with an emergency room visit increases during (HR = 1.57, 95% CI: 0.99, 2.47) and after (HR = 1.80, 95% CI = 1.13, 2.87) well development, compared to before well development. With inclusion of unexposed control patients in the CITS analyses this association attenuates towards the null (HR = 1.11, 95% CI: 0.79, 1.56) or after (HR = 1.24, 95% CI: 0.90, 1.70) O&G well site development.

Stratified analyses indicates that age modifies the risk of an AF event with an emergency room visits during O&G well development (Table 3). In at risk patients aged >80 years, ITS and CITS results for the during well development period are similar and indicate that risk of an AF event increases during well development. In the during well development period, risk of an AF event with an emergency room visit doubled (HR = 2.55, 95% CI: 1.50, 4.36) in at-risk patients aged >80 years. The results indicate that the risk does not persist past the well development period and show no increased risk in younger patients. Stratified analysis did not indicate biological sex, co-morbidities, or geographical region as effect modifiers (results not shown).

Sensitivity analyses excluding patients: for whom we could not confirm that the street address did not change over our follow-up period (Supplementary Material Table S6), with well development durations within the 25th to 75th percentile range (Supplementary Material Table S7); and living at an elevation less than 6,000 feet (Supplementary Material Table S8) did not inferentially change our results.

Discussion

Our results provide strong evidence (47) that older AF patients living within 0.39 miles (2,059 feet) of an O&G well site may experience increased AF exacerbation during site development with the possibility of increased AF exacerbation risk up to at least 0.8 miles (4,224) feet from the site, which does not persist past the well development period. Our results also suggest that AF patients identified as female living within 0.39 miles (2,059 feet) of an O&G site may experience increased AF exacerbation

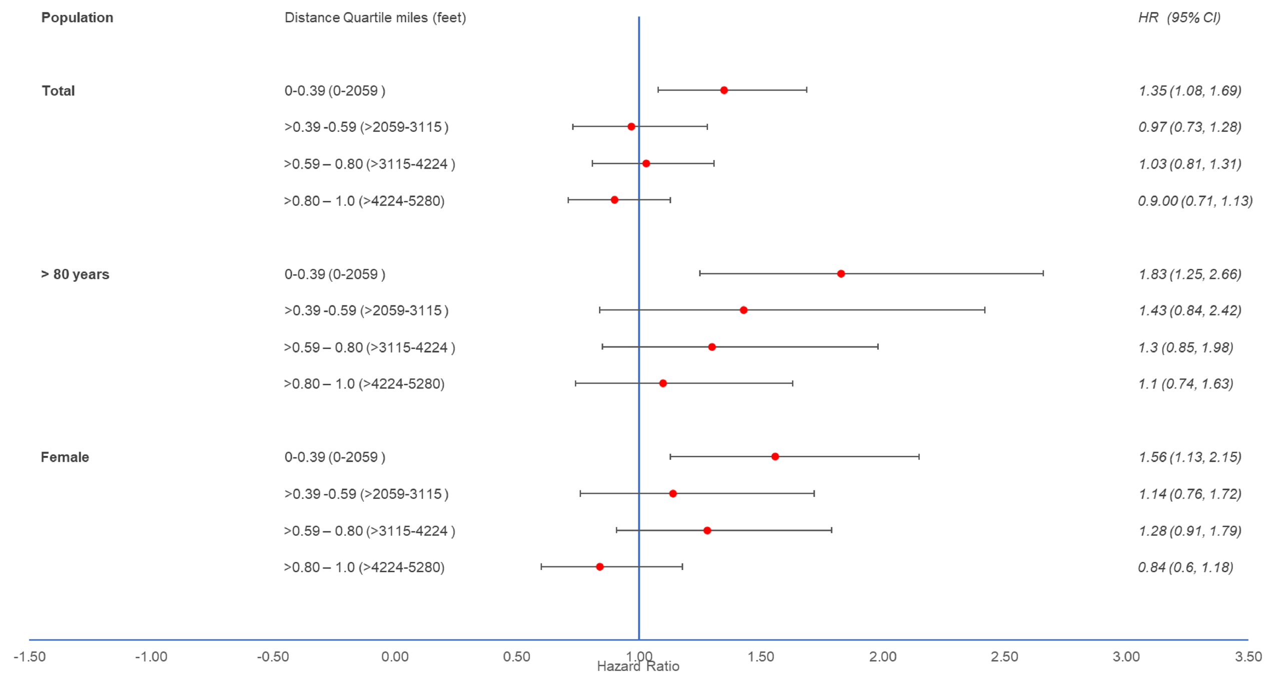


FIGURE 3
Hazard ratios with 95% confidence intervals from multi-failure survival analysis with control for total population, patients aged >80 years, and patients identified as female by distance quartile from oil and gas well development site: HRs of at-risk to controls for an atrial fibrillation event during O&G well site development compared to before development. HR, hazard ratio; CI, confidence interval.

TABLE 3 Results from multi-failure survival analysis for AF event with an emergency room visit: hazard ratios for an AF event during and after well development periods compared to before well development.

Analysis	Interrupted time series analysis[a]		Parallel analysis (exposed vs. unexposed in before period)[b]	Controlled interrupted time series	
	HR during[b] (95% CI)	HR after[a] (95% CI)	HR (95% CI)	HR during[c] (95% CI)	HR after[c] (95% CI)
Total population (main analysis)	1.57 (0.99, 2.47)	1.80 (1.13, 2.87)	0.81 (0.56, 1.19)	1.11 (0.79, 1.56)	1.24 (0.90, 1.70)
>80 years	2.67 (1.26, 5.64)	1.1 (0.60, 2.01)	0.95 (0.47, 1.92)	2.55 (1.50, 4.36)	1.10 (0.60, 2.01)
74–80 years	1.20 (0.37, 3.93)	2.40 (0.82, 6.99)	0.56 (0.23, 1.40)	0.58 (0.25, 1.32)	1.34 (0.71, 2.53)
66–73 years	1.43 (0.54, 3.75)	2.71 (1.09, 6.78)	0.58 (0.27, 1.26)	0.76 (0.39, 1.47)	1.48 (0.82, 2.67)
<66 years	0.78 (0.31, 1.96)	1.11 (0.43, 2.85)	1.36 (0.66, 2.78)	0.77 (0.36, 1.64)	1.04 (0.52, 2.09)
Females	1.93 (1.0, 3.74)	1.64 (0.80, 3.37)	0.68 (0.40, 1.17)	1.30 (0.83, 2.03)	0.97 (0.59, 1.61)
Males	1.25 (0.67, 2.35)	1.94 (1.05, 3.59)	0.98 (0.58, 1.68)	0.97 (0.58, 1.61)	1.57 (1.04, 2.36)

AF, atrial fibrillation; CI, confidence interval; HR, hazard ratio.
[a]Does not include control patients.
[b]Adjusted for sex, age at first AF event, elevation of residence, duration of well development, hypertension, diabetes, and region, and exposure status.
[c]Adjusted for sex, age at first AF event, elevation of residence, duration of well development, hypertension, diabetes, and region, interaction between period (before, during, after development) and exposure status.

during site development, which does not appear to persist past the well development period. We did not observe increases in AF exacerbation in younger AF or male patients. Previous studies indicating that people living near O&G well sites may experience alterations in vascular function associated with AF (45), heart failure exacerbation (43), and increased hospitalization for acute MI (44), as well as exposures to noise and air pollution levels known to affect cardiovascular health (36) support these results. These important and biologically plausible findings contribute further epidemiological evidence that environmental stressors exacerbate AF.

Air and noise pollution emitted during the development of O&G well sites potentially impact all individuals residing in the vicinity of the sites (32). Exposure to noise elicits an acute stress reaction characterized by autonomic nervous system response, specifically, increased sympathetic activity (58), which plays an important role in the initiation and maintenance of AF (59). On the molecular level, beta adrenergic stimulation triggers an intracellular signaling cascade that can lead to intracellular calcium overload creating a particularly arrhythmogenic environment that promotes triggered activity. Resulting depolarizations generate spontaneous ectopy. Simultaneously, enhanced automaticity, promoted by increased circulating catecholamines, also leads to focal ectopic atrial activity. Both triggered activity and enhanced automaticity are believed to be the primary drivers for AF initiation. This stress reaction has been observed in response to road traffic noise (60); thus, it is plausible that exposure to stressful noise levels may induce AF in susceptible individuals. Additionally, alterations in autonomic tone, inflammation, oxidative stress, and changes in intracardiac filling pressures are known triggers for AF (58, 61–68) and are reported in response to $PM_{2.5}$ exposure (15, 18, 69–75). Exposure to $PM_{2.5}$ has been associated with increased blood pressure and acute alteration in vascular function, which may contribute to hypertension, an AF risk factor (1, 73, 76–79).

Interestingly, our results indicate that living near development of an O&G well site has a greater impact on older and female AF patients. Other studies also have observed that older adults living in close proximity to O&G well sites may bear greater health and mortality risks than younger adults (43, 80). Additionally, prior studies report that both women and the elderly are at higher risk of mortality and CV mortality when exposed to elevated $PM_{2.5}$ levels (81). Our findings may be explained by age- and gender-related changes in response to physiologic stressors. Significantly higher levels of cortisol have been observed in women compared to age-matched men and older vs. younger subjects when exposed to psychological or cognitive challenges (82). It is plausible that older subjects spend more time at home, increasing the duration of exposure (83).

Our observation that AF exacerbation risk does not to persist past the well development period indicates that the increased risk is transitory in nature. A transitory increase in AF exacerbation risk could worsen AF patient acute outcomes, as evidenced by the increased risk for AF claims associated with an emergency room visit.

Our study benefited from an efficient design that accounts for unmeasured confounding, accurate definition of before, during, and after O&G well site development periods, and the availability of sequential measures of AF diagnoses and related morbidities in the COAPCD. Additionally, our temporal control design features allowed us to account for risk factors that drive AF development in an accumulating manner and time-varying variables such as season and regional air pollution events (e.g., wildfires).

Nonetheless, our study had some limitations. While our CITS design allowed us to account static environmental stressors and time varying environmental stressors a the regional level, it did not account for changing environmental stressors a the local level that may have occurred during the follow up period,, such as construction activities and development of other O&G well sites further than the closet site,. This may have biased result towards or away from the null. Assuming the street address in the COAPCD is the residential address and the possibility for change of residence in our study cohort may introduce exposure misclassification. However, our sensitivity analysis on patients for whom we could confirm the street address over the follow up period indicates that exposure misclassification from change in residential address had little effect on our results (Supplementary Material Table S6). It also is possible that some AF and co-morbidity claims were misdiagnosed. Our claim review confirmed that most (91%) patients had plausible AF diagnoses, with similar results for at-risk and control patients. Thus, this is mostly a concern for null results. Not all AF incidents may result in an COAPCD claim and not all AF patients are included in the COAPCD. Therefore, our results, may underrepresent the true incidence of AF. This too is mostly a concern for null results. It is important to appreciate that our outcome is an AF claim in the COAPCD and not new onset AF. Therefore, our results apply to the prevalence of AF. Because we did not include AF patients with addresses that could not be geocoded, our results may not be generalizable to the whole Colorado AF patient population. Because noise and air pollution measures were not available for this retrospective study, we could not elucidate specific associations between noise or air pollution and AF. Because the COAPCD includes only 65% of Colorado's population, our results may not represent the 35% of Coloradans that are uninsured or privately insured.

The prospect that proximity to O&G well site development, a significant noise and air pollution source, increases AF exacerbation risk requires attention. Health care providers should be aware of the increased risk for AF during O&G well site development for their older and female patients and provide appropriate patient education to help mitigate risk. Additionally, these findings support development of mitigation strategies and regulations to protect the health of populations living near O&G well sites. While this study advances understanding on relationships between residential proximity to development of O&G well sites and AF exacerbation, a future prospective cohort study that can follow populations for AF over the course of O&G well site development will be necessary to understand the etiological relationships between specific environmental stressors, such as noise and air pollution, and incidence and severity of AF events.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: The dataset was obtained from the Colorado All Payer Claims Dataset thru a data use agreement with the Center for Improving Value in Health Care. A data use agreement with Center for Improving Value in Health Care is required to access the data. Requests to access these datasets should be directed to Eddy Costa: ECosta@CIVHC.org.

Ethics statement

The studies involving humans were approved by Colorado Multi-Institutional Review Board, University of Colorado. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

LM: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. WA: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. BA: Conceptualization, Formal Analysis, Funding acquisition, Validation, Writing – original draft, Writing – review & editing. CT: Conceptualization, Funding acquisition, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article.

Research reported in this publication was supported by the National Heart, Lung, And Blood Institute of the National Institutes of Health under Award Number R21HL145337 and the Center for Improving Value in Health Care (CIVHC). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or CIVHC.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fepid.2024.1379271/full#supplementary-material

References

1. Lip GYH, Fauchier L, Freedman SB, Van Gelder I, Natale A, Gianni C, et al. Atrial fibrillation. *Nat Rev Dis Primers*. (2016) 2:16016. doi: 10.1038/nrdp.2016.16

2. Williams BA, Honushefsky AM, Berger PB. Temporal trends in the incidence, prevalence, and survival of patients with atrial fibrillation from 2004 to 2016. *Am J Cardiol*. (2017) 120(11):1961–5. doi: 10.1016/j.amjcard.2017.08.014

3. Schnabel RB, Yin X, Gona P, Larson MG, Beiser AS, McManus DD, et al. 50 year trends in atrial fibrillation prevalence, incidence, risk factors, and mortality in the framingham heart study: a cohort study. *Lancet*. (2015) 386(9989):154–62. doi: 10.1016/S0140-6736(14)61774-8

4. Colilla S, Crow A, Petkun W, Singer DE, Simon T, Liu X. Estimates of current and future incidence and prevalence of atrial fibrillation in the U. S. adult population. *Am J Cardiol*. (2013) 112(8):1142–7. doi: 10.1016/j.amjcard.2013.05.063

5. Jackson SL, Tong X, Yin X, George MG, Ritchey MD. Emergency department, hospital inpatient, and mortality burden of atrial fibrillation in the United States, 2006 to 2014. *Am J Cardiol*. (2017) 120(11):1966–73. doi: 10.1016/j.amjcard.2017.08.017

6. Centers for Disease Control and Prevention. Atrial Fibrillation Fact Sheet. Available online at: https://www.cdc.gov/dhdsp/data_statistics/fact_sheets/fs_atrial_fibrillation.htm (accessed January 31, 2018).

7. Shao Q, Liu T, Korantzopoulos P, Zhang Z, Zhao J, Li G. Association between air pollution and development of atrial fibrillation: a meta-analysis of observational studies. *Heart Lung*. (2016) 45(6):557–62. doi: 10.1016/j.hrtlng.2016.08.001

8. Monrad M, Sajadieh A, Christensen JS, Ketzel M, Raaschou-Nielsen O, Tjønneland A, et al. Long-term exposure to traffic-related air pollution and risk of incident atrial fibrillation: a cohort study. *Environ Health Perspect*. (2017) 125 (3):422–7. doi: 10.1289/EHP392

9. Monrad M, Sajadieh A, Christensen JS, Ketzel M, Raaschou-Nielsen O, Tjonneland A, et al. Residential exposure to traffic noise and risk of incident atrial fibrillation: a cohort study. *Environ Int*. (2016) 92-93:457–63. doi: 10.1016/j.envint.2016.04.039

10. Andrade J, Khairy P, Dobrev D, Stanley N. The clinical profile and pathophysiology of atrial fibrillation: relationships among clinical features, epidemiology, and mechanisms. *Circ Res*. (2014) 114(9):1453–8. doi: 10.1161/CIRCRESAHA.114.303211

11. Xue X, Hu J, Xiang D, Li H, Jiang Y, Fang W, et al. Hourly air pollution exposure and the onset of symptomatic arrhythmia: an individual-level case-crossover study in 322 Chinese cities. *CMAJ*. (2023) 195(17):E601–e611. doi: 10.1503/cmaj.220929

12. Liao D, Shaffer ML, He F, Rodriguez-Colon S, Wu R, Whitsel EA, et al. Fine particulate air pollution is associated with higher vulnerability to atrial fibrillation–the APACR study. *J Toxicol Environ Health A*. (2011) 74(11):693–705. doi: 10.1080/15287394.2011.556056

13. Link MS, Luttmann-Gibson H, Schwartz J, Mittleman MA, Wessler B, Gold DR, et al. Acute exposure to air pollution triggers atrial fibrillation. *J Am Coll Cardiol*. (2013) 62(9):816–25. doi: 10.1016/j.jacc.2013.05.043

14. Rich DQ, Mittleman MA, Link MS, Schwartz J, Luttmann-Gibson H, Catalano PJ, et al. Increased risk of paroxysmal atrial fibrillation episodes associated with acute increases in ambient air pollution. *Environ Health Perspect*. (2006) 114(1):120–3. doi: 10.1289/ehp.8371

15. Rich DQ, Freudenberger RS, Ohman-Strickland P, Cho Y, Kipen HM. Right heart pressure increases after acute increases in ambient particulate concentration. *Environ Health Perspect*. (2008) 116(9):1167–71. doi: 10.1289/ehp.11230

16. Solimini AG, Renzi M. Association between air pollution and emergency room visits for atrial fibrillation. *Int J Environ Res Public Health*. (2017) 14(6):661. doi: 10.3390/ijerph14060661

17. Vencloviene J, Babarskiene RM, Dobozinskas P, Dedele A, Lopatiene K, Ragaisyte N. The short-term associations of weather and air pollution with emergency ambulance calls for paroxysmal atrial fibrillation. *Environ Sci Pollut Res Int*. (2017) 24(17):15031–43. doi: 10.1007/s11356-017-9138-7

18. Buteau S, Goldberg MS. A structured review of panel studies used to investigate associations between ambient air pollution and heart rate variability. *Environ Res*. (2016) 148:207–47. doi: 10.1016/j.envres.2016.03.013

19. Thacher JD, Poulsen AH, Hvidtfeldt UA, Raaschou-Nielsen O, Ketzel M, Jensen SS, et al. Long-term exposure to transportation noise and risk for atrial fibrillation: a danish nationwide cohort study. *Environ Res*. (2022) 207:112167. doi: 10.1016/j.envres.2021.112167

20. Dahlquist M, Frykman V, Stafoggia M, Qvarnström E, Wellenius GA, Ljungman PLS. Short-term ambient air pollution exposure and risk of atrial fibrillation in patients with intracardiac devices. *Environ Epidemiol*. (2022) 6(4):e215. doi: 10.1097/EE9.0000000000000215

21. Dahlquist M, Frykman V, Kemp-Gudmunsdottir K, Svennberg E, Wellenius GA, Ljungman PLS. Short-term associations between ambient air pollution and acute atrial fibrillation episodes. *Environ Int*. (2020) 141:105765. doi: 10.1016/j.envint.2020.105765

22. Sadlonova M, Senges J, Nagel J, Celano C, Klasen-Max C, Borggrefe M, et al. Symptom severity and health-related quality of life in patients with atrial fibrillation: findings from the observational ARENA study. *J Clin Med*. (2022) 11(4):1140. doi: 10.3390/jcm11041140

23. Münzel T, Gori T, Babisch W, Basner M. Cardiovascular effects of environmental noise exposure. *Eur Heart J*. (2014) 35(13):829–36. doi: 10.1093/eurheartj/ehu030

24. Jarup L, Babisch W, Houthuijs D, Pershagen G, Katsouyanni K, Cadum E, et al. Hypertension and exposure to noise near airports: the HYENA study. *Environ Health Perspect*. (2008) 116(3):329–33. doi: 10.1289/ehp.10775

25. Pirrera S, De Valck E, Cluydts R. Nocturnal road traffic noise: a review on its assessment and consequences on sleep and health. *Environ Int*. (2010) 36(5):492–8. doi: 10.1016/j.envint.2010.03.007

26. Carrington MJ, Trinder J. Blood pressure and heart rate during continuous experimental sleep fragmentation in healthy adults. *Sleep*. (2008) 31(12):1701–12. doi: 10.1093/sleep/31.12.1701

27. Wang Q, Xi B, Liu M, Zhang Y, Fu M. Short sleep duration is associated with hypertension risk among adults: a systematic review and meta-analysis. *Hypertens Res*. (2012) 35:1012. doi: 10.1038/hr.2012.91

28. Sabanayagam C, Shankar A. Sleep duration and cardiovascular disease: results from the national health interview survey. *Sleep*. (2010) 33(8):1037–42. doi: 10.1093/sleep/33.8.1037

29. United States Energy Information Adminstration. Drilling Productivity Report. (2017).

30. Czolowski ED, Santoro RL, Srebotnjak T, Shonkoff SB. Toward consistent methodology to quantify populations in proximity to oil and gas development: a national spatial analysis and review. *Environ Health Perspect*. (2017) 125(8):086004. doi: 10.1289/EHP1535

31. McKenzie LM, Allshouse WB, Burke T, Blair BD, Adgate JL. Population size, growth, and environmental justice near oil and gas wells in Colorado. *Environ Sci Technol*. (2016) 50(21):11471–80. doi: 10.1021/acs.est.6b04391

32. Adgate JL, Goldstein BD, McKenzie LM. Potential public health hazards, exposures and health effects from unconventional natural gas development. *Environ Sci Technol*. (2014) 48(15):8307–20. doi: 10.1021/es404621d

33. King GE. *Hydraulic Fracturing 101: What Every Representative, Environmentalist, Regulator, Reporter, Investor, University Researcher, Neighbor and Engineer Should Know About Estimating Frac Risk and Improving Frac Performance in Unconventional Gas and Oil Wells*. Woodlands, TX: Semantics Scholar (2012).

34. Hecobian A, Clements AL, Shonkwiler KB, Zhou Y, MacDonald LP, Hilliard N, et al. Air toxics and other volatile organic compound emissions from unconventional oil and gas development. *Environ Sci Technol Lett*. (2019) 6(12):720–6. doi: 10.1021/acs.estlett.9b00591

35. Radtke C, Autenrieth DA, Lipsey T, Brazile WJ. Noise characterization of oil and gas operations. *J Occup Environ Hyg*. (2017) 14(8):659–67. doi: 10.1080/15459624.2017.1316386

36. Allshouse WB, McKenzie LM, Barton K, Brindley S, Adgate JL. Community noise and air pollution exposure during the development of a multi-well oil and gas pad. *Environ Sci Technol*. (2019) 53(12):7126–85. doi: 10.1021/acs.est.9b00052

37. Witter RZ, McKenzie L, Stinson KE, Scott K, Newman LS, Adgate J. The use of health impact assessment for a community undergoing natural gas development. *Am J Public Health*. (2013) 103(6):1002–10. doi: 10.2105/AJPH.2012.301017

38. McKenzie LM, Guo R, Witter RZ, Satvitz DA, Newman LS, Adgate JL. Birth outcomes and maternal residential proximity to natural gas development in rural Colorado. *Environ Health Perspect*. (2014) 122:412–7. doi: 10.1289/ehp.1306722

39. McKenzie LM, Allshouse W, Daniels S. Congenital heart defects and intensity of oil and gas well site activities in early pregnancy. *Environ Int*. (2019) 132:104949. doi: 10.1016/j.envint.2019.104949

40. Janitz AE, Dao HD, Campbell JE, Stoner JA, Peck JD. Association between benzene and congenital anomalies in Oklahoma, 1997–2009. *Occup Environ Med*. (2018) 75(11):822–9. doi: 10.1136/oemed-2018-105054

41. Tang IW, Langlois PH, Vieira VM. Birth defects and unconventional natural gas developments in Texas, 1999–2011. *Environ Res*. (2021) 194:110511. doi: 10.1016/j.envres.2020.110511

42. Willis MD, Carozza SE, Hystad P. Congenital anomalies associated with oil and gas development and resource extraction: a population-based retrospective cohort study in Texas. *J Expo Sci Environ Epidemiol*. (2023) 33(1):84–93. doi: 10.1038/s41370-022-00505-x

43. McAlexander TP, Bandeen-Roche K, Buckley JP, Pollak J, Michos ED, McEvoy JW, et al. Unconventional natural gas development and hospitalization for heart failure in Pennsylvania. *J Am Coll Cardiol*. (2020) 76(24):2862–74. doi: 10.1016/j.jacc.2020.10.023

44. Denham A, Willis MD, Croft DP, Liu L, Hill EL. Acute myocardial infarction associated with unconventional natural gas development: a natural experiment. *Environ Res*. (2021) 195:110872. doi: 10.1016/j.envres.2021.110872

45. McKenzie LM, Crooks J, Peel JL, Blair BD, Brindley S, Allshouse WB, et al. Relationships between indicators of cardiovascular disease and intensity of oil and natural gas activity in northeastern Colorado. *Environ Res*. (2019) 170:56–64. doi: 10.1016/j.envres.2018.12.004

46. Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int J Epidemiol*. (2017) 46(1):348–55. doi: 10.1093/ije/dyw098

47. Lopez Bernal J, Cummins S, Gasparrini A. The use of controls in interrupted time series studies of public health interventions. *Int J Epidemiol*. (2018) 47(6):2082–93. doi: 10.1093/ije/dyy135

48. McKenzie LM, Blair B, Hughes J, Allshouse WB, Blake NJ, Helmig D, et al. Ambient nonmethane hydrocarbon levels along Colorado's northern front range: acute and chronic health risks. *Environ Sci Technol*. (2018) 52(8):4514–25. doi: 10.1021/acs.est.7b05983

49. Holder C, Hader J, Avanasi R, Hong T, Carr E, Mendez B, et al. Evaluating potential human health risks from modeled inhalation exposures to volatile organic compounds emitted from oil and gas operations. *J Air Waste Manage Assoc*. (2019) 69:1503–24. doi: 10.1080/10962247.2019.1680459

50. Weisner ML, Allshouse WB, Erjavac BW, Valdez AP, Vahling JL, McKenzie LM. Health symptoms and proximity to active multi-well unconventional oil and gas development sites in the city and county of broomfield, Colorado. *Int J Environ Res Public Health*. (2023) 20(3):2634. doi: 10.3390/ijerph20032634

51. Deziel NC, Brokovich E, Grotto I, Clark CJ, Barnett-Itzhaki Z, Broday D, et al. Unconventional oil and gas development and health outcomes: a scoping review of the epidemiological research. *Environ Res*. (2020) 182:109124. doi: 10.1016/j.envres.2020.109124

52. Commission COaGC. Colorado Oil and Gas Information System. Available online at: https://cogcc.state.co.us/cogis/ (accessed June 30, 2014).

53. Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J Am Stat Assoc*. (1989) 84(408):1065–73. doi: 10.1080/01621459.1989.10478873

54. Allison PD. *Survival Analysis Using SAS: A Practical Guide Second Edition*. 2nd ed. Cary, NC USA: SAS Institute Inc. (2010).

55. Lin W, Zhu T, Xue T, Peng W, Brunekreef B, Gehring U, et al. Association between changes in exposure to air pollution and biomarkers of oxidative stress in children before and during the Beijing olympics. *Am J Epidemiol*. (2015) 181(8):575–83. doi: 10.1093/aje/kwu327

56. Nianogo RA, Benmarhnia T, O'Neill S. A comparison of quasi-experimental methods with data before and after an intervention: an introduction for epidemiologists and a simulation study. *Int J Epidemiol*. (2023) 52(5):1522–33. doi: 10.1093/ije/dyad032

57. Wasserstein RL, Lazar NA. The ASA statement on p-values: context, process, and purpose. *Am Stat*. (2016) 70(2):129–33. doi: 10.1080/00031305.2016.1154108

58. Ulrich-Lai YM, Herman JP. Neural regulation of endocrine and autonomic stress responses. *Nat Rev Neurosci*. (2009) 10(6):397–409. doi: 10.1038/nrn2647

59. Chen P-S, Chen LS, Fishbein MC, Lin S-F, Nattel S. Role of the autonomic nervous system in atrial fibrillation. *Pathophysiol Ther*. (2014) 114(9):1500–15. doi: 10.1161/CIRCRESAHA.114.303772

60. Babisch W. The noise/stress concept, risk assessment and research needs. *Noise and Health*. (2003) 5(18):1–11.

61. Aviles RJ, Martin DO, Apperson-Hansen C, Houghtaling PL, Rautaharju P, Kronmal RA, et al. Inflammation as a risk factor for atrial fibrillation. *Circulation*. (2003) 108(24):3006–10. doi: 10.1161/01.CIR.0000103131.70301.4F

62. Hu Y-F, Chen Y-J, Lin Y-J, Chen S-A. Inflammation and the pathogenesis of atrial fibrillation. *Nat Rev Cardiol*. (2015) 12:230. doi: 10.1038/nrcardio.2015.2

63. Otto CM. Heartbeat: inflammatory biomarkers in atrial fibrillation. *Heart*. (2016) 102(7):485–6. doi: 10.1136/heartjnl-2016-309537

64. Gutierrez A, Van Wagoner DR. Oxidant and inflammatory mechanisms and targeted therapy in AF: an update. *J Cardiovasc Pharmacol*. (2015) 66(6):523–9. doi: 10.1097/FJC.0000000000000313

65. Li J, Solus J, Chen Q, Rho YH, Milne G, Stein CM, et al. The role of inflammation and oxidative stress in atrial fibrillation. *Heart Rhythm*. (2010) 7(4):438–44. doi: 10.1016/j.hrthm.2009.12.009

66. Van Wagoner DR. Oxidative stress and inflammation in atrial fibrillation: role in pathogenesis and potential as a therapeutic target. *J Cardiovasc Pharmacol*. (2008) 52(4):306–13. doi: 10.1097/FJC.0b013e31817f9398

67. Samman Tahhan A, Sandesara PB, Hayek SS, Alkhoder A, Chivukula K, Hammadah M, et al. Association between oxidative stress and atrial fibrillation. *Heart Rhythm*. (2017) 14(12):1849–55. doi: 10.1016/j.hrthm.2017.07.028

68. Lam CSP, Rienstra M, Tay WT, Liu LCY, Hummel YM, van der Meer P, et al. Atrial fibrillation in heart failure with preserved ejection fraction: association with exercise capacity, left ventricular filling pressures, natriuretic peptides, and left atrial volume. *JACC Heart Fail*. (2017) 5(2):92–8. doi: 10.1016/j.jchf.2016.10.005

69. Fiordelisi A, Piscitelli P, Trimarco B, Coscioni E, Iaccarino G, Sorriento D. The mechanisms of air pollution and particulate matter in cardiovascular diseases. *Heart Fail Rev*. (2017) 22(3):337–47. doi: 10.1007/s10741-017-9606-7

70. Cole-Hunter T, Weichenthal S, Kubesch N, Foraster M, Carrasco-Turigas G, Bouso L, et al. Impact of traffic-related air pollution on acute changes in cardiac autonomic modulation during rest and physical activity: a cross-over study. *J Expo Sci Environ Epidemiol*. (2015) 26:133. doi: 10.1038/jes.2015.66

71. Peters A, Fröhlich M, Döring A, Immervoll T, Wichmann HE, Hutchinson WL, et al. Particulate air pollution is associated with an acute phase response in men. Results from the MONICA–Augsburg study. *Eur Heart J*. (2001) 22(14):1198–204. doi: 10.1053/euhj.2000.2483

72. Kelly FJ, Fussell JC. Role of oxidative stress in cardiovascular disease outcomes following exposure to ambient air pollution. *Free Radic Biol Med*. (2017) 110:345–67. doi: 10.1016/j.freeradbiomed.2017.06.019

73. Brook RD, Brook JR, Urch B, Vincent R, Rajagopalan S, Silverman F. Inhalation of fine particulate air pollution and ozone causes acute arterial vasoconstriction in healthy adults. *Circulation*. (2002) 105(13):1534–6. doi: 10.1161/01.CIR.0000013838.94747.64

74. Møller P, Christophersen DV, Jacobsen NR, Skovmand A, Gouveia ACD, Andersen MHG, et al. Atherosclerosis and vasomotor dysfunction in arteries of animals after exposure to combustion-derived particulate matter or nanomaterials. *Crit Rev Toxicol*. (2016) 46(5):437–76. doi: 10.3109/10408444.2016.1149451

75. Cascio WE. Proposed pathophysiologic framework to explain some excess cardiovascular death associated with ambient air particle pollution: insights for public health translation. *Biochim Biophys Acta*. (2016) 1860(12):2869–79. doi: 10.1016/j.bbagen.2016.07.016

76. Brook RD, Rajagopalan S. Particulate matter, air pollution, and blood pressure. *J Am Soc Hypertens*. (2009) 3(5):332–50. doi: 10.1016/j.jash.2009.08.005

77. Dales R, Liu L, Szyszkowicz M, Dalipaj M, Willey J, Kulka R, et al. Particulate air pollution and vascular reactivity: the bus stop study. *Int Arch Occup Environ Health*. (2007) 81(2):159–64. doi: 10.1007/s00420-007-0199-7

78. Rundell KW, Hoffman JR, Caviston R, Bulbulian R, Hollenbach AM. Inhalation of ultrafine and fine particulate matter disrupts systemic vascular function. *Inhal Toxicol*. (2007) 19(2):133–40. doi: 10.1080/08958370601051727

79. Nurnberger J, Keflioglu-Scheiber A, Opazo Saez AM, Wenzel RR, Philipp T, Schafers RF. Augmentation index is associated with cardiovascular risk. *J Hypertens*. (2002) 20(12):2407–14. doi: 10.1097/00004872-200212000-00020

80. Li L, Dominici F, Blomberg AJ, Bargagli-Stoffi FJ, Schwartz JD, Coull BA, et al. Exposure to unconventional oil and gas development and all-cause mortality in medicare beneficiaries. *Nat Energy*. (2022) 7(2):177–85. doi: 10.1038/s41560-021-00970-y

81. Franklin M, Zeka A, Schwartz J. Association between PM2.5 and all-cause and specific-cause mortality in 27 US communities. *J Expo Sci Environ Epidemiol*. (2007) 17(3):279–87. doi: 10.1038/sj.jes.7500530

82. Seeman TE, Singer B, Wilkinson CW, McEwen B. Gender differences in age-related changes in HPA axis reactivity. *Psychoneuroendocrinology*. (2001) 26(3):225–40. doi: 10.1016/S0306-4530(00)00043-3

83. Otte C, Hart S, Neylan TC, Marmar CR, Yaffe K, Mohr DC. A meta-analysis of cortisol response to challenge in human aging: importance of gender. *Psychoneuroendocrinology*. (2005) 30(1):80–91. doi: 10.1016/j.psyneuen.2004.06.002

ATTACHMENT C

STUDY 18

Journal of the Air & Waste Management Association

ISSN: 1096-2247 (Print) 2162-2906 (Online) Journal homepage: www.tandfonline.com/journals/uawm20

Evaluation of gas well setback policy in the Marcellus Shale region of Pennsylvania in relation to emissions of fine particulate matter

Zoya Banan & Jeremy M. Gernand

To cite this article: Zoya Banan & Jeremy M. Gernand (2018) Evaluation of gas well setback policy in the Marcellus Shale region of Pennsylvania in relation to emissions of fine particulate matter, Journal of the Air & Waste Management Association, 68:9, 988-1000, DOI: 10.1080/10962247.2018.1462866

To link to this article: https://doi.org/10.1080/10962247.2018.1462866

JOURNAL OF THE AIR & WASTE MANAGEMENT ASSOCIATION
2018, VOL. 68, NO. 9, 988–1000
https://doi.org/10.1080/10962247.2018.1462866

Taylor & Francis
Taylor & Francis Group

TECHNICAL PAPER

Evaluation of gas well setback policy in the Marcellus Shale region of Pennsylvania in relation to emissions of fine particulate matter

Zoya Banan and Jeremy M. Gernand

Department of Energy and Mineral Engineering, Pennsylvania State University, University Park, PA

ABSTRACT
Shale gas has become an important strategic energy source with considerable potential economic benefits and the potential to reduce greenhouse gas emissions in so far as it displaces coal use. However, there still exist environmental health risks caused by emissions from exploration and production activities. In the United States, states and localities have set different minimum setback policies to reduce the health risks corresponding to the emissions from these locations, but it is unclear whether these policies are sufficient. This study uses a Gaussian plume model to evaluate the probability of exposure exceedance from EPA concentration limits for PM2.5 at various locations around a generic wellsite in the Marcellus shale region. A set of meteorological data monitored at ten different stations across Marcellus shale gas region in Pennsylvania during 2015 serves as an input to this model. Results indicate that even though the current setback distance policy in Pennsylvania (500 ft. or 152.4 m) might be effective in some cases, exposure limit exceedance occurs frequently at this distance with higher than average emission rates and/or greater number of wells per wellpad. Setback distances should be 736 m to ensure compliance with the daily average concentration of PM2.5, and a function of the number of wells to comply with the annual average PM2.5 exposure standard.

Implications: The Marcellus Shale gas is known as a significant source of criteria pollutants and studies show that the current setback distance in Pennsylvania is not adequate to protect the residents from exceeding the established limits. Even an effective setback distance to meet the annual exposure limit may not be adequate to meet the daily limit. The probability of exceeding the annual limit increases with number of wells per site. We use a probabilistic dispersion model to introduce a technical basis to select appropriate setback distances.

PAPER HISTORY
Received October 4, 2017
Revised March 14, 2018
Accepted April 4, 2018

Introduction

During the past decade, shale gas development has become more economical due to recent technological achievements (Jacoby, O'Sullivan, and Paltsev 2011). Many consider natural gas to be a bridging fuel toward a cleaner energy system, which allows the electrical generation system to continue using fossil-based infrastructures and helps to reduce greenhouse gas emissions through the displacement of coal use. Shale gas can also provide an improvement in public and occupational health, and can reduce average environmental impacts from energy production as it replaces coal-produced electricity (Jenner and Lamadrid 2013). The United States holds large reserves of shale gas, so exploitation of this resource is expected to continue for many decades. Some of the most famous reserves are the Barnett Shale in Texas, the Denver–Julesberg Basin in Colorado, and the Marcellus Shale in the northeast. Shale gas production in Pennsylvania started in 2007 and increased to more than 4 trillion cubic feet in 2014 (EIA—Shale gas production 2016). According to the Energy Information Administration (EIA), Pennsylvania possesses 56.2 trillion cubic feet shale gas proved reserves in 2014 (EIA—Shale gas proved reserves 2016). Thus, continued exploitation is expected.

However, shale gas exploration activities can influence local air quality. While vertical drilling is usually enough to get to the conventional gas reservoirs, shale gas development requires a combination of vertical and horizontal drilling that adds up to a length considerably longer than a conventional wellbore. Also shale oil and gas development needs hydraulic fracturing by means of high-pressure fluids to create fractures down the wellbore and into the target rock so that the oil or gas flows out (Ogoke et al. 2014; Vidic et al. 2013). Thus, shale gas development causes a larger number of

CONTACT Jeremy M. Gernand jmgernand@psu.edu Department of Energy and Mineral Engineering, The Pennsylvania State University, 110 Hosler Bldg., University Park, PA 16802.
Supplemental data for this paper can be accessed on the publisher's website.

engines to run over a longer period of time. Application of large diesel-powered equipment or gas turbines during exploring (i.e., drilling and hydraulic fracturing) stages and also the use of diesel trucks for transportation can affect the air quality within the vicinity of the well site and even farther downwind. Even though the emissions from shale gas production can be offset by the decrease in the emissions due to replacing fuels like coal by natural gas at the end use (Pacsi et al. 2013), these emissions can cause severe health issues within local areas around shale gas development sites.

This analysis presented here would apply to oil wells as well if it required a similar operation time. However, in the Marcellus region, unconventional gas wells predominate. Coal mines and other permanent or long-term energy-related installations require more extensive review of their environmental impacts by federal, state, and local authorities than short-term drilling and fracturing operations that last only a few weeks. It is our implication in this paper that the temporary effects due to gas or oil drilling with hydraulic fracturing may exceed expectations and the mitigating effects of the existing setback policy may not be sufficient in some cases.

Emissions from shale gas activities are mainly characterized to be volatile organic compounds (VOCs), nitrogen dioxide, sulfur dioxide, and particulate matter (Shonkoff, Hays, and Finkel 2014; Zielinska, Fujita, and Campbell 2011). These pollutants can cause acute diseases, such as respiratory symptoms, lung and heart diseases, and chronic health impacts, such as cancer (Adgate, Goldstein, and McKenzie 2014; Kelly and Fussell 2012). Therefore, public concern exists regarding hazardous air pollutants (HAPs) associated with unconventional gas development activities (Olaguer 2012). The U.S. Environmental Protection Agency (EPA) has set National Ambient Air Quality Standards (NAAQS) that regulate standards on concentrations of criteria pollutants, namely, carbon monoxide (CO), lead (Pb), ozone (O_3), particulate matter (PM), nitrogen dioxide (NO_2), and sulfur dioxide (SO_2). While development of each gas well might require a relatively short period of time, and the environmental health effect would be expected to be small, the unprecedented expansion of activity in regions such as the Marcellus Shale, with thousands of new wells drilled each year, could mean that the impact is more significant than would be accounted for with a single-well analysis.

On the other hand, different states have set setback policies in order to reduce the corresponding health risks to people due to emission concentrations higher than standard within the vicinity of shale gas well sites. Setback policy regulates the minimum distance required between occupied buildings or occupied outdoor areas and the site of the gas well. Nevertheless, Fry (2013) finds that there was no technical basis in the designation of setback distances in 26 municipalities in the Dallas–Fort Worth Metroplex, Texas. Most of these setback distances are set through a compromise among governments, the regulated community, environmental and citizen interest groups, and landowners (Haley et al. 2016). According to the section 3215 of 2015 Pennsylvania Consolidated Statutes (58 PA Cons Stat § 3215), the existing setback limit from residential buildings is 152.4 m (500 ft) in case of an unconventional gas well. Haley et al. (2016) investigates the sufficiency of current setback distances in Texas, Pennsylvania, and Colorado using VOCs emission measurements taken by others. Based on these evaluations, the authors suggest that the current setback requirement in the Marcellus Shale of Pennsylvania is not sufficient to maintain human exposure below the established limits for benzene and hydrogen sulfide (Haley et al. 2016).

A few extant studies are dedicated to modeling the dispersion of hazardous pollutants originated from oil and gas activities (Olaguer 2012; Rodriguez, Barna, and Moore 2009). Olaguer (2012) simulated ozone concentration using an average wind speed (4.8 m/sec) and direction (southwest) at 1 p.m. CST at Fort Worth, TX, in June 2011 by means of an Eulerian air quality model. However, as such results correspond to specific wind speed and direction from the source, they are not qualified to be generalized to all cases (i.e., different well site patterns and different locations within the same distance as the evaluated ones). Therefore, this serves as a limitation in evaluating the current setback policies.

Other relevant studies conducted for shale gas areas are mainly focused on modeling ozone, VOCs, and NO_x dispersion, and very few are aimed at modeling PM concentration (Olaguer 2012; Rodriguez, Barna, and Moore 2009). Rodriguez, Barna, and Moore (2009) evaluated changes in concentration of ozone originated due to oil and gas development in the western United States within a 36-km grid using the Eulerian dispersion model CAMx. This study used CAMx as a chemistry transport model to simulate ozone formed through chemical reaction of NO_x and VOCs. Results of Rodriguez, Barna, and Moore (2009) indicate that based on background level of ozone concentration in the western United States, reported between 40 and 70 ppb, ozone concentration level caused from oil and gas activities might lead to exceedance of the EPA ozone standard. However, to the best of our knowledge, current literature lacks a robust modeling of PM emission distribution associated with shale gas activities.

Different modeling tools have been used for dispersion simulation at different scale. Touma et al. (2006)

introduced Eulerian dispersion models as a grid-based regional scale tool that is capable of treating transport and chemical transformation of air toxics. These models are suitable for modeling the formation and transport of ozone, acid rain, and PM. However, the article argues that these models are not appropriate for simulating the air toxics with local impact when finer spatial resolutions are required. Air toxics modeling of pollutant emissions can be demonstrated at four spatial scales: national, regional, urban, and local. National-scale estimates mainly aim to characterize the average risk across the country that the general population might face and to give a better picture of the toxic air problem. On the other hand, local-scale models, such as the approach this paper presents, can help demonstrate concentration level and exposure risks very close to specific sources or within their neighborhood. Also, some studies investigate emission level changes at regional scale, such as Roy, Adams, and Robinson (2014), which simulates regional $PM_{2.5}$, NO_x, and VOCs emission rates from the Marcellus region under 2009 conditions and for the case of emission control technologies application.

One main and critical input to such dispersion modeling tools is known to be the emission rate. Emissions during the shale gas development process mainly originate from diesel engines (Roy, Adams, and Robinson 2014), and therefore, evaluation methods are designed based on analysis of these engines. In 1972, EPA published a list of emission factors, required for developing air pollutants emission inventories, in an online document titled AP-42 (Compilation of Air Pollutant Emissions Factors 1972). Efforts have been made to amend the listed emission factors in AP-42. Shah et al. (2004) did several on-road measurements and laboratory analysis of samples from diesel engines to make an estimate of PM, elemental carbon (EC), and organic carbon (OC) emissions from these engines. Using the factors provided by AP-42, Roy, Adams, and Robinson (2014) developed an emission inventory of NOx, VOCs, and $PM_{2.5}$ from major activities in Marcellus Shale gas regions specifically located in Pennsylvania and portions of West Virginia and New York.

The goal of this study is to evaluate the minimum necessary distance from a Pennsylvania shale gas well site to avoid local exceedance of the air quality standards for particulate matter, considering the variety of the numbers of wells per site in addition to variable emissions rates during drilling and hydraulic fracturing. By employing an emissions dispersion model across the range of meteorological conditions expected in Pennsylvania for any future Marcellus shale gas well, this study calculates the probability of exceeding EPA NAAQS for $PM_{2.5}$ at various distances and directions from a generic well site, and compares these results to the current setback policy.

Methodology

Data sources

Wind data comprising wind direction, wind speed, and relative humidity, measured at 10 weather monitoring stations in Pennsylvania all through the year 2015, served as an input to the emissions dispersion model. These stations are Altoona–Blair County Airport, Allegheny County Airport, Bedford Regional Airport, DuBois Regional Airport, Erie International Airport, Port Meadville Airport, Johnstown–Cambria County Airport, Pocono Mountains Municipal Airport, Penn Valley Airport, and Pittsburgh International Airport. These stations are located in the areas where Marcellus Shale gas development activities have occurred since year 2000. The wind data were accessed through Iowa Environmental Mesonet (IEM 2016). IEM reports wind data for every 20 min at specific locations. Here, we used only one wind speed and direction measurement per hour, based on the mean values if multiple measurements were available or based on the only existing measurement for each hour if measurements were missing.

For the purpose of this study, we define a generic well site that would represent any well site in the Marcellus Shale region of Pennsylvania. We generate multiple cases so that each of them considers the well site to be located closer to one of the monitoring stations. For each case, we used measured wind data at the closest monitoring station to model the emission dispersion from development of the well site.

EPA's latest NAAQS (National Ambient Air Quality Standards 2012) set the annual primary and secondary standard levels for $PM_{2.5}$ as 12 $\mu g/m^3$ and 15 $\mu g/m^3$, respectively, and the daily standard for both of them to be equal to 35 $\mu g/m^3$. According to EPA, primary standards provide public health protection while secondary standards provide public welfare protection.

For the purpose of this analysis, we base our estimation of the $PM_{2.5}$ emission rate at well sites on estimated $PM_{2.5}$ emission rate by Roy, Adams, and Robinson (2014) over 1 year for each well. This study estimated the mean and 95% confidence interval for $PM_{2.5}$ emission rate from drilling and hydraulic fracturing of one shale gas well to be equal to 0.3 (0.03—1) tons/yr well drilled and 0.16 (0.03–0.4) tons/yr well drilled, respectively. Roy, Adams, and Robinson (2014) estimated emission rates based on emission factors reported by

EPA's inventory models (AP-42) and other literature for diesel engines with size similar to that of Marcellus drill rig engines and fracturing pumps. Also they performed a Monte Carlo approach to quantify the emission factors and other variables of the emission equations (e.g., engines' horsepower, load factor, number of fracking stages, etc.) using each variable specific distribution; for variables with rich data sets, such as emission factors, Roy, Adams, and Robinson (2014) took advantage of distribution of actual data, but triangular and uniform distribution was used in case of an input with limited data sets.

We estimated the hourly rate of $PM_{2.5}$ emission from one single well using the reported 95% confidence interval by Roy, Adams, and Robinson (2014). First, according to interviews with unconventional gas development experts and also the discussion by Ogoke et al. (2014), we set the time frame at 14 days for drilling and 9 days for hydraulic fracturing of one shale gas well. Then, based on these time periods and annual emission rates by Roy, Adams, and Robinson (2014), we calculate the 95% interval for hourly rate of $PM_{2.5}$ emission during drilling and hydraulic fracturing of one well to be 0.81 (0.09—2.7) kg/hr and 0.67 (0.14—1.68) kg/hr, respectively. Last, we applied the values corresponding to mean and high (97.5th percentile value) emission rate levels in order to generate an overview of the concentrations and also to give an estimate of a likely and conservative considerations regarding limitation of health risks.

Use of a median value for emission rate would be a better choice to estimate the most likely emissions, as the mean is influenced more by extremes and outliers. However, there are limited data on emission rate measurement, and Monte Carlo results by Roy, Adams, and Robinson (2014) do not include the exact value for the median. From the cumulative distribution functions provided by Roy, Adams, and Robinson (2014), the median value for emission rate is approximately 0.54 kg/hr for both drilling and hydro fracking, which is equal to 67% and 81% of the respective mean values. Sensitivity analysis on the effect of displacement of mean emission rate level by the median on necessary minimum distances to meet the standards is discussed in the supplemental file to this paper.

Generally, particulate matter is known to be made up of a number of components, including acids (such as nitrates and sulfates), organic chemicals, metals, and soil or dust particles (World Health Organization 2003). To estimate the composition of emissions associated with shale gas activities, many reports on local emission analysis and also on composition of emission from different types of sources were reviewed (Corbett and Abruzzo 2014; EPA National Emissions Inventory [NEI] 2014; Zielinska, Fujita, and Campbell 2011). This study considers particulate matter composition to be 45% elemental carbon (EC), 35% nitric acids, and 20% ammonium nitrates. Sensitivity analysis on the effect of $PM_{2.5}$ composition on necessary minimum distances to meet the standards is discussed in the supplemental file to this paper.

A Gaussian plume model assumes that no chemical reaction occurs with the dispersed particles involved. However, the increase in relative humidity causes the particle size to increase by a factor that depends on the dry particle size, particle type, and also level of humidity (Gopoch, Burk, and Davidson 1980; Martin and Finlay 2005; Sinclair, Countess, and Hoopes 1974; Winkler 1988). For example, Popovicheva et al. (2008a) shows that based on the hydrophobic or hydrophilic nature of soot particles, one particle can uptake one to eight monolayers of water on its surface. For particles in the form of aqueous droplets, studies took advantage of Kohler theory to estimate the changes to particles diameter as a result of interaction with water (Akpootu and Gana 2013; Petters and Kreidenweis 2007). As a result of change in the particle size, relative humidity affects aerosol concentration (Gopoch, Burk, and Davidson 1980). Since elemental carbon has a smaller molecular weight than the other types of $PM_{2.5}$ particles, the influence of water uptake through adsorption and absorption (Popovicheva et al. 2008b) on the particle concentration can be more significant.

Model

Air pollution models are powerful tools to quantify the relationship between emission rate and changes in ambient concentration. As it is not feasible to measure pollutant concentration at every single location, these models are becoming more indispensable for regulatory and research applications. Touma et al. (2006) discussed two major types of air quality models, namely, local-scale (source-based) dispersion models and a regional-scale (grid-based) chemical transport model. For the purpose of this research, the simulation method is a Gaussian plume model, the basic method used to estimate concentration in local-scale models. Our model treats the shale gas well site as a point source of emission and simulates the dispersion of emissions from development activities for every hour. Thus, it allows for probabilistic evaluation of concentration exceedance of EPA NAAQS through consideration of all possible time periods and multiple locations. The output of this model is a probability map of concentrations, rather than a concentration map. Also, our code makes it easier to track the trend of concentration changes downwind of specific wind directions and perform the sensitivity analysis on different variables and inputs, such as wind data and emission composition.

The $PM_{2.5}$ emissions predominately come from diesel engines located at the site. Fugitive dust, mineral dust from proppant handling, and emissions from delivery trucks and related vehicles have not been included. While the most accurate representation is that these diesel engines are an array of point sources located close to the center of the well site, rather than a single point source, the configuration of this array and its relationship to the specific location of the well are not consistent from site to site. Since the effect of interest is the concentration of $PM_{2.5}$ at distances greater than 152.4 m (500 ft) from the well location, we consider it a reasonable simplifying assumption to locate all emissions at the center of the activity.

Gaussian plume model is a governing advection–diffusion equation, mainly used over a short range (within 50 km), describing the movement of pollutants in the atmosphere. This model uses the average wind characteristic data (speed and direction) over a specific period of time and its output is an average estimation of the pollutants concentration at specific location(s).

Gaussian plume equation is as shown in eq 1:

$$C(x,y,z,t) = \frac{Q}{2\pi u\sigma_y\sigma_z}\, exp\left(-\frac{y^2}{2\sigma_y^2}\right)\left[exp\left(-\frac{(z-H_{eff})^2}{2\sigma_z^2}\right) + exp\left(-\frac{(z+H_{eff})^2}{2\sigma_z^2}\right)\right] \tag{1}$$

where C is the substance concentration as a function of x, y, z, and time (t), x is the distance downwind from the stack, y is the crosswind distance from the plume centerline, z is the vertical distance from the ground level, Q is the source emission rate, u is the average wind speed at stack height, σ_y is dispersion coefficient in the crosswind direction, σ_z is dispersion coefficient in the vertical direction, and H_{eff} is the effective stack height.

To calculate σ_y and σ_z values, the Briggs' formula tabulated by Arystanbekova (2004) is used. Roughly, stability class of moderately unstable is defined for simulations associated to weather data measured each day between 5 am to 8 pm. Stability class of slightly stable considered in generating simulation result based on weather data measured between 8 pm to 5 am daily. Stability is a function of wind velocity and sky cloudiness (Arystanbekova 2004). However, there was limited type of data on the sky cloudiness that could serve as an input to the model. We defined the stability based on distribution of monitored wind velocity values at the mentioned weather stations during day and night hours. Based on the wind data, almost 75% of the measured wind velocity values lie within the range of 2 to 5 (m/s) wherein choosing stability status of moderately unstable during day and slightly stable during night sounds reasonable. About 20% of the hourly measured wind velocities were beyond 5 (m/s) that in average implies neutral stability class during day and night (Arystanbekova 2004). In case of neutral stability the calculated downwind concentration would be higher versus moderately unstable status and lower versus slightly stable status. As the model output is an average of concentrations at all hours, consideration of the same stability as the ones defined for 75% of the hours for the hours wherein wind velocities were beyond 5 (m/s) would be a reasonable approximation in overall.

The assumption of a zero stack height is conservative. Emissions of $PM_{2.5}$ from shale gas development mainly originate from diesel engines (Roy, Adams, and Robinson 2014), which are located at the ground level. These engines have a variety of configurations, and do not all emit exhaust at the same height. In this study H_{eff} is set to zero. Sensitivity analysis on the effect of H_{eff} on necessary minimum distances to meet the standards is discussed in the supplemental file to this paper.

The presence of any structures around the emission source could affect the concentrations in the near-field. Some modeling tools try to treat the downwash due to presence of buildings and other structures, but there seem to be overpredictions and underpredictions involved (Peterson and Beyer-Lout 2012; Peterson, Guerra, and Bova 2017). However, shale gas development activities typically occur in the rural areas where the probability of existence of such structures within very close vicinity of a well site is low. Thus, as the goal of this work is to provide a generic evaluation of shale gas development effect on the local air quality, we assume that there is no such structure within the vicinity of the generic well site.

In this study, the Gaussian plume model was implemented with MATLAB to simulate PM concentrations. The original model code, developed by Paul Connolly from the University of Manchester (Connolly 2014), used eq 1 with consideration of one constant value for wind speed and direction. We have modified this code in order to consider the role of time in emission dispersion procedure by taking hourly wind speed and direction measurements as an input for any period of time, instead of a single value. The modified implementation considers the emission characterization (aerosols size, molecular mass, density, etc.), the effect of humidity on aerosols concentration, and the change in atmosphere stability at different hour of the day. The code simulates all possible cases and provides the probabilistic evaluation of cases wherein exceedance from concentration standards occur.

Analysis

In this study, we assess emission concentrations to determine the minimum distance from the source that is required for an occupied area to be located in order to not experience any exceedance from $PM_{2.5}$ concentration standards. To ensure that conditions across the Marcellus region were represented, wind data from 10 selected monitoring stations in the state of Pennsylvania are used. These stations are selected as to be located close to Marcellus Shale gas development areas in Pennsylvania. Bootstrapping was used to complete the wind data for missing hourly measurements.

Even though Gaussian plume models may be an appropriate modeling tool for long distances (typically within 50 km from the source) (Touma et al. 2006), wind profiles might change over this distance. However, this analysis did not extend the calculated dispersion beyond 5 km. In this steady-state model, it is assumed that wind speed and direction are constant within the vicinity of the well site where concentrations are modeled on an hourly basis. Moreover, we set the goal of this work to demonstrate just the role of shale gas development on the quality of the ambient air within vicinity of the well sites. Thus, the calculated concentrations in the model only originate from drilling and hydraulic fracturing activities at the well site, and background concentrations are not considered. This assumption also implies that no accumulation of emissions is presumed from hour to hour.

Background emission concentrations are those generated from other natural and anthropogenic sources such as motor vehicles on the road, factories, and other distant emission sources. EPA provides $PM_{2.5}$ summary data reports for individual monitoring sites at different counties and cities in Pennsylvania (EPA 2017a, 2017b). According to these air quality reports, background concentration at different locations in Pennsylvania in 2015 took annual average values within the range of 5.9 to 13.8 μg/m^3 with a mean of 10.3 μg/m^3 and daily average values within the range of 0 to 63.5 μg/m^3 with a mean of 10.3 μg/m^3. However, these values are not available at the locations within the vicinity of most of the developed shale gas well sites.

Even though the background concentration has an important impact on defining the setback distance in more polluted areas, consideration of an average value introduces more uncertainty into the results due to underestimation or overestimation at different locations. Therefore, the model output becomes a less comprehensive representative of the changes in the air quality due to shale gas activities.

The model produces hourly $PM_{2.5}$ concentrations at all locations within the vicinity of a representative shale gas well site over the drilling time period. The appropriate time frame is defined based on the number of wells on the well pad. Using the wind data from 10 stations available for every hour during January 1, 2015, to December 31, 2015, the code generates arrays of PM concentrations indicating the locations where exceedance of EPA NAAQS occurs on a probabilistic basis.

Exceedance plots are generated based on two time-averaging approaches: annual and daily concentration averages. To calculate the percentage of exceedance occurrence based on annual average concentration, first, all the possible drilling time periods during a year are identified based on number of wells per well site. Then, using every set of wind data (from the 10 weather monitoring stations), the annual average concentration is modeled for all the plausible time periods during a year. Thus, the percentage of exceedance is defined to be equal to the percentage of the times that the annual average concentration at each direction exceeds the annual standard, i.e., 12 μg/m^3 for $PM_{2.5}$.

To plot the safe area boundary based on daily average approach, first the average aerosol concentration is calculated using the wind data for every 24 hr. Then, the percentage of exceedance is defined to be equal to the percentage of days that daily average concentration at each direction exceeds the daily standard, i.e., 35 μg/m^3 for $PM_{2.5}$.

By definition, compliance with EPA's $PM_{2.5}$ annual standard of 12 μg/m^3 is calculated by averaging the annual mean concentrations over three consecutive years. Also, compliance with EPA's $PM_{2.5}$ daily standard of 35 μg/m^3 is determined by calculating the 98th percentile of all 365 daily averages each year, and then averaging together three successive years' 98th percentiles. However, shale gas development well sites are temporary point sources of emissions that usually exist for significantly less than a 3-year period of time. We set the purpose of this work to perform an evaluation that is favorable from a public health point of view and to provide a critique of current policy, instead of establishing violations of existing regulations, to avoid any exceedance experience of annual and daily standards of $PM_{2.5}$. For the rest of this paper, we investigate any case wherein one year average concentration exceeds the annual standard and any case wherein one day average concentration exceeds the daily standard.

Results

Annual average concentration of $PM_{2.5}$ emissions was modeled at radial distances from the representative well site using the wind data records in 2015. Figure 1 depicts boundaries of areas where concentrations exceed the annual standard 5% and 0% of the times based on annual averaging for two cases of well pad comprising of one well and six wells. Results are demonstrated for the two emission rate levels, mean (0.81 kg/hr for drilling and 0.67 for hydro fracking) and high (2.7 kg/hr for drilling and 1.68 for hydro fracking). The current Pennsylvania's residential setback distance (500 ft or 152.4 m) from the shale gas well is displayed by the red-dashed circle. To calculate the annual average concentrations, one to six wells per well site were considered based on the permit records from Marcellus Shale gas development in Pennsylvania (Department of Oil and Gas Reporting website 2016). Figure 2 depicts the histogram of number of permitted wells per well pad in Marcellus Shale region of Pennsylvania in 2015.

While exceedance of the annual standard is unlikely to occur at the current setback distance in case of a well pad with a single well, the probability of exceeding this standard increases with a greater number of wells. For example, a typical well pad comprising of six wells can cause exceedance occurrence at a specific location with respect to the well site even at mean expected emission rate. Figure 1c shows that for a typical well pad consisting of six wells, a residential area must be located at least 67–158 m away from the center of the well pad, depending on the compass direction, to be certain of no exceedance of annual standard at mean emission rate. This distance range increases to 121–291 m at the high emission rate (Figure 1d). The 95% confidence intervals for the minimum safe distance range from a well pad with six wells are 62–137 m (Figure 1c) and 113–248 m (Figure 1d) at mean and high emission rates, respectively.

As a clarification, for instance, to generate the plot presented in Figure 1a, the time period to develop one well is equal to 23 days: 14 days for drilling and 9 days for hydraulic fracturing. For all possible 23-day contiguous time periods during the year 2015, we collected wind data from each of the 10 available measurement sites. For each time period, we modeled concentrations within the vicinity of a generic well site and set concentrations on remaining days to 0 in 2015, and calculated the annual average concentration for that specific case. We repeated these calculations for all the plausible time periods. At the last step, we identified the

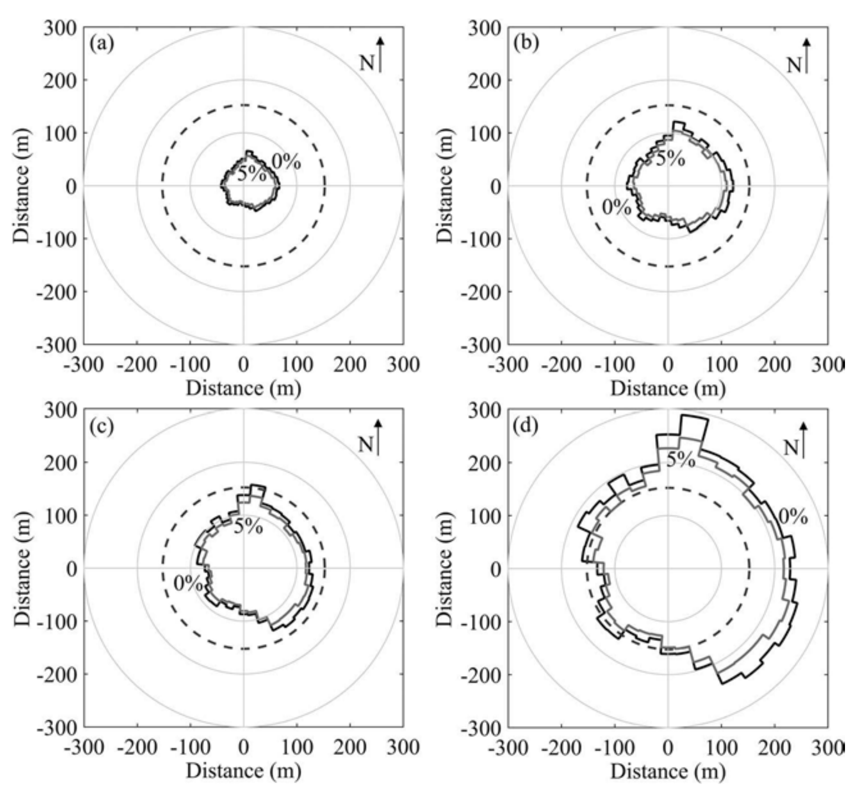


Figure 1. Distance from well pad to maintain safe level of concentration based on EPA's $PM_{2.5}$ annual concentration standard, for the cases: (a) one well at mean emission rate, (b) one well at high emission rate, (c) six wells at mean emission rate, and (d) six wells at high emission rate. The emission source is assumed to be located at the origin. The dashed circle indicates the locations at the current Pennsylvania setback limit (500 ft or 152.4 m) from the source.

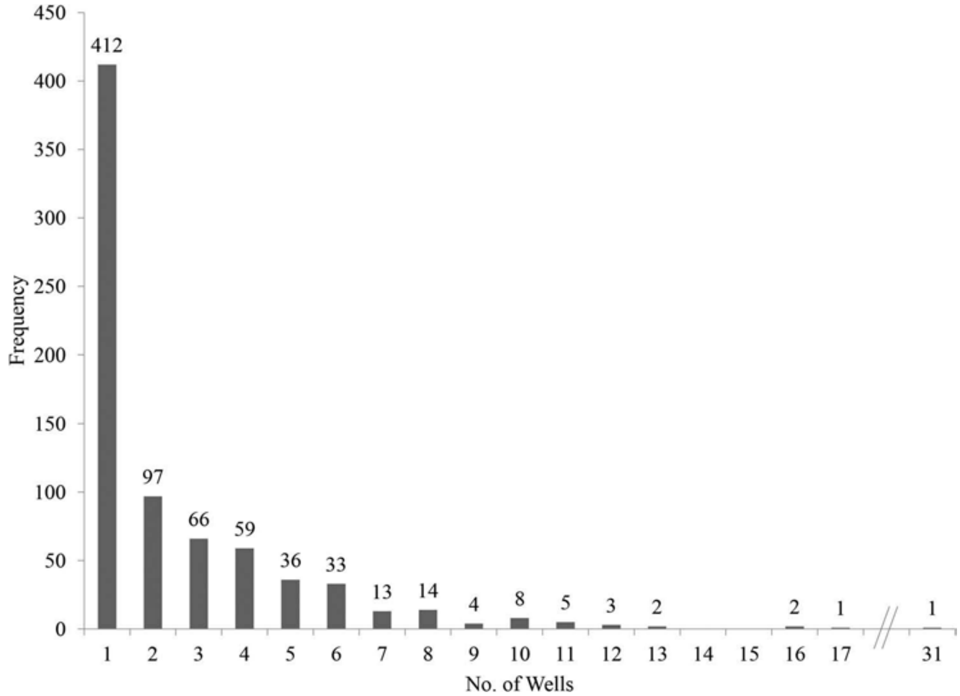


Figure 2. Histogram of number of permitted wells per well pad in Marcellus shale region of Pennsylvania in 2015; 93% of the well pads contain 6 wells or fewer.

locations where exceedance of EPA's annual concentration standard occurs in 5% of the sets of results.

These results indicate the effect of changes in number of wells and emission rates on the minimum residential distance required for no exceedance with the probability of higher than 5%. According to Figure 1, the south and southeast wind directions are the ones that imply the farthest safe distances from the source. Figure 3 presents the trend of change in safe distance values versus number of wells corresponding to the south wind direction for two levels of emission rate.

Results indicate that in case of one well per well site, occupied areas should be located no closer than about 67 m away at mean emission rate and about 122 m away at the high emission rate in order to not experience any concentration above EPA's annual standard. Thus, Pennsylvania's setback distance seems to be effective for these cases. However, these distances are a

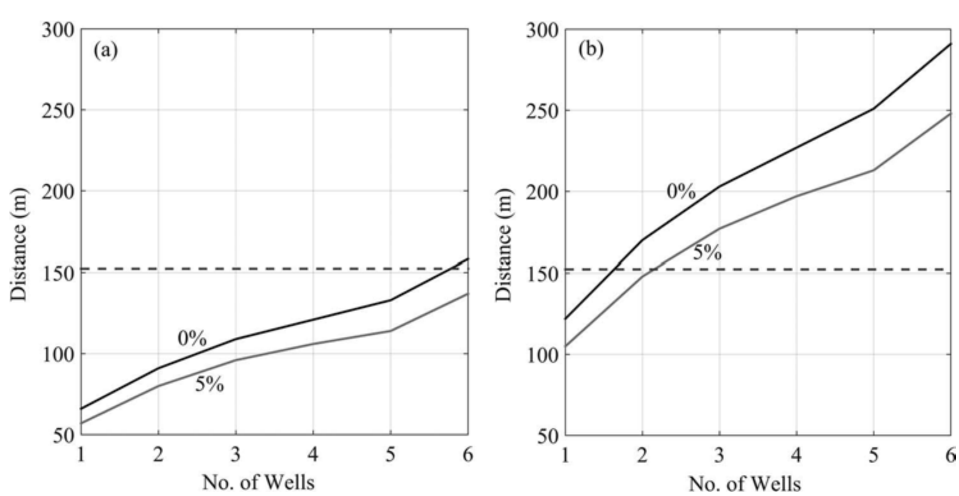


Figure 3. Minimum distance versus number of wells to meet annual concentration standard at the north of the source (south wind direction) for the cases: (a) mean emission rate, and (b) high emission rate. The dashed line indicates the current Pennsylvania setback limit (500 ft or 152.4 m).

function of the number of wells per site (more wells means longer drilling and fracturing periods), and the necessary distances are found to be equal to at least 158 m at mean emission rate and about 291 m at high emission rate in the case of six wells per well site.

Figure 4 demonstrates boundaries of areas where exceedance of daily standard on $PM_{2.5}$ occurs 5% and 0% of the times. Results are demonstrated for the two emission rate levels, mean and high. Again, the current Pennsylvania residential setback distance (500 ft or 152.4 m) from the source is presented by the red-dashed circle. Figure 4 demonstrates that in order to meet the daily standard, residences must be at least 272–371 m away from the generic well pad, depending on the compass direction, for the mean emission rate. This distance range increases to 530–736 m for the high emission rate. In order to not experience any concentration exceedance more than 5% of the time, the corresponding minimum distance requirements from the generic well pad are 101–208 m and 189–407 m for mean and high emission rates, respectively.

The simulations indicate that the minimum distance of at least 371 m in case of mean emission rate and the minimum distance of at least 736 m in case of high emission rate is required in order to be certain of no exceedance occurrence of the daily standard.

Results from this simulation indicate that at mean emission rate, the highest probability of concentration exceedance at 152.4 m (500 ft) from annual limit is 3% for the case of a well pad with six wells. This value increases to 87% at high emission rate for the same number of wells. The locations that these percentage values represent are reported in Table 1.

Discussion

While Roy, Adams, and Robinson (2014) discuss the regional contribution of $PM_{2.5}$ emissions alongside NOx and VOCs originated from Marcellus development in Pennsylvania, they indicate a relatively moderate $PM_{2.5}$ contribution when averaged across the region. These results, however, help to shed light on the more significant, though heterogeneous, local effects that occur at specific locations in the vicinity of well sites.

Arguments by Haley et al. (2016), based on their evaluation of current setbacks efficiency, support the fact that at the current setback distance in Pennsylvania, people are not protected from potential health effects of VOC

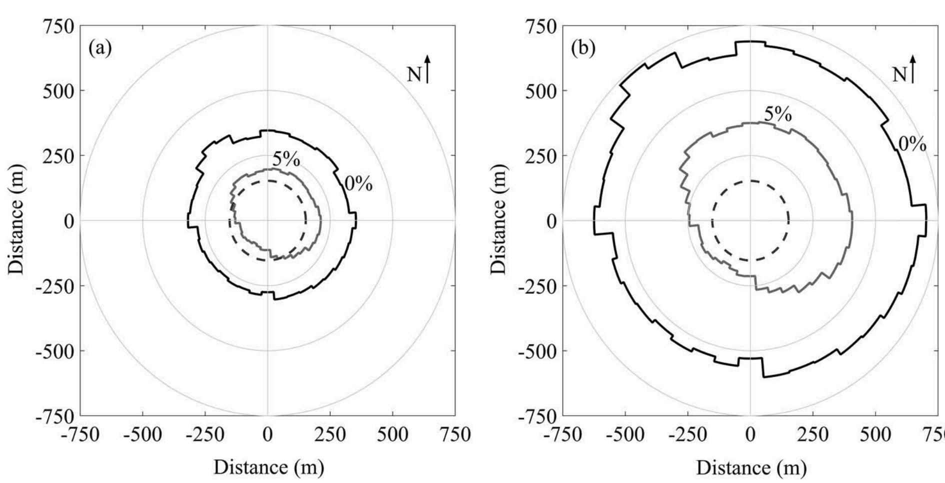


Figure 4. Distance from well pad to maintain safe level of concentration based on EPA's $PM_{2.5}$ daily concentration standard, for the cases: (a) mean emission rate, and (b) high emission rate. The emission source is assumed to be located at the origin. The dashed circle indicates the locations at the current Pennsylvania setback limit (500 ft or 152.4 m) distance from the source.

Table 1. Probability of concentration exceedance of annual concentration standard at 152.4 m (500 ft) in case of a well pad with six wells.

	Mean emission rate		High emission rate	
Probability levels of concentration exceedance	Percentage (%)	Wind direction	Percentage (%)	Wind direction
Lowest	0	All except for south to south-southwest	0	North to north-northeast East-northeast to east
Highest	3	South to south-southwest	87	South to south-southwest

Notes. Values indicate the direction and the percentage of the time that concentrations exceed the annual standard at that direction for mean and high emission rate, at the current setback limit of 152.4 m (500 ft); e.g., there is a residential located on the "S–SSW" wind direction path that experiences the annual standard exceedance 87% of the time.

emissions. Similarly to Haley et al. (2016), the results from this analysis imply that the current Pennsylvania setback limit for natural gas wells is not sufficient to ensure inhabited areas meet the EPA's $PM_{2.5}$ daily standard and it is not sufficient to guarantee no exceedance of annual standard for sites with multiple wells per pad. A minimum distance of at least 736 m (about 2400 ft) is required in order to ensure concentrations less than EPA's daily average $PM_{2.5}$ standard. Even at this distance, there is still a slight chance that exceedance of daily standard occurs. The emission rates used in this simulation process are the ones reported by Roy, Adams, and Robinson (2014) as the 95% confidence interval for $PM_{2.5}$ emission rate. Therefore, even at the suggested setback distance, there is still a probability of 2.5% of experiencing emission rates that would cause exceedance of concentration standards.

Roohani et al. (2017) predicts the regional ozone and $PM_{2.5}$ concentration using the modeling tool CAMx over a 36 km × 36 km grid resolution under three different scenarios that are defined based on three levels of shale gas development activities in 2020. Results by Roohani et al. (2017) demonstrate a relatively small change in the mean annual $PM_{2.5}$ concentration due to shale gas activities under the three scenarios at regional scale (0.1–0.4 $\mu g/m^3$). However, our results find these changes to be more significant at local scale as a result of different densities of shale gas activities. For instance, we find the increase in the mean annual $PM_{2.5}$ concentration of a residential located at 152.4 m distance to the north of a well site comprising 6 wells to get up to about 20 and 60 $\mu g/m^3$ at mean and high emission rates, respectively.

The current Pennsylvania setback limit is in fact sufficient to protect occupied areas from exceedance of the annual standard assuming mean emission rates. However, the required distance in this case is a function of the number of wells located at the site. Setback distances from natural gas development should take the density of that development into account, as currently known regulations use the same distance regardless of the number of wells—the number of wells being a proxy for the length of time that high intensity activities will be occurring at the site. Given that the main risks from $PM_{2.5}$ exposure are chronic diseases such as cancer and heart disease (Lepeule et al. 2012), it would be prudent to treat the concentrations above the average annual concentration standard as a higher priority. Even assuming average emission rates, the current setback policy is insufficient for sites with more than five wells, and for high emission rates, the current setback distance is only sufficient for a single well per pad.

A limit of no more than one well per well pad would serve to ensure no exceedances of the annual average $PM_{2.5}$ standard occurred assuming 95th percentile emissions of $PM_{2.5}$. Review of permit data sets through 2001–2015 (Department of Oil and Gas Reporting website 2016) shows that a significant percentage of developed well pads have had more than one well per pad (e.g., 7% of sites in 2015 had more than 6 wells; see Figure 2). A limit on the number of wells per pad should be accompanied by study on the economic and environmental trade-offs required, but such information is not currently available.

The increase of setback distances for natural gas drilling would likely make some parts of the Marcellus Shale inaccessible to gas recovery, at least temporarily. However, as technology continues to stretch the maximum lateral lengths possible, this may not remain a restriction. As drilling costs related to increasing lateral lengths are proprietary, it is not possible to evaluate the impact of increasing setback distance on them. In Pennsylvania, most new drilling activities occur in sparsely populated areas of the commonwealth, so it may be possible to adopt increased setback distances without significant impact, especially considering the fact that they would be temporary restrictions. However, this is unlikely to be the case in more densely populated areas like Allegheny County, which contains the city of Pittsburgh and related suburban communities. The economic effects of such a change are not expected to be exclusively negative, however, as increasing distance from a well has been associated with increasing property values (Boxall, Chan, and McMillan 2005).

Application of increased setback distance standards may not be quite sufficient by itself to provide human health protection (Haley et al. 2016). There are alternative policies to consider in lieu of increasing setback distances. It would be possible to maintain the current setback distance of 152.4 m (500 ft) in Pennsylvania if policymakers set a cap on the $PM_{2.5}$ emissions rate from these sites at 0.165 kg/hr. This value represents an emission rate of only 20% and 25% of the mean emission rate used in this analysis for drilling and hydraulic fracturing, respectively. Such a standard might seem stringent, but it would negate the need for a 480% increase in the setback distance to prevent exceedance of the daily average $PM_{2.5}$ standard. As some well services companies are increasing their use of gas turbines to provide power, rather than diesel engines, such a reduction in PM emissions may be possible.

This analysis addresses possible exceedances of the concentration standards for $PM_{2.5}$, though several other pollutants of interest are emitted during gas exploration and production, such as nitrogen oxides, ozone, VOCs, sulfur oxides, and PM_{10}. $PM_{2.5}$ is one of the most significant quantities emitted during these activities, and the concentration standards have been established relatively recently based on current health risk research. Continuing study should examine setback policy in light of each of these pollutants.

While the simplifying assumptions for this analysis including a constant emissions rate are reasonable for prediction of a generic future gas well, these assumptions would not necessarily be applicable for the determination of health risks accepted by current policy for past wells. In reality, emission rates likely differ based on depth and length of the well, so it may be possible to estimate in advance whether or not the emission rate at a particular site would be high, average, or low, and to evaluate the risk and the necessary setback distance on that basis.

These results are based on the assumption that future distributions of wind speed and direction will remain consistent with those recorded in 2015. The wind speed and direction around specific well sites are dependent not just on the overall distribution of weather patterns, but also on the specific geography of the site, including hills, trees, ridges, and so on. Some modeling tools apply some modifications in order to consider the influence of these complexities on the model output. For example, AERMOD takes the base elevation and hill height scale data as input and considers their influence in modeling the dispersion of a plume (EPA 2016). However, each of these features can increase or decrease the concentrations near the wells, depending on the specifics, and setback policy as a useful heuristic in place of doing extensive modeling of each well site should be based on a generic or flat terrain in order to be applicable to different cases and locations. Also precipitation is not included in this model, and would be expected to increase the settling rate of fine particulate matter, thus reducing the concentrations on those days with rain or snow. However, these results provide expected $PM_{2.5}$ concentrations on dry days, and while the overall probability of an exceedance might change with the inclusion of precipitation, the 0% exceedance distance presented here would not change.

The concentrations predicted by our model at the location of EPA's monitoring stations are below the measurements at these locations (i.e., most of the recorded emissions at these sites come from other sources), but these stations are located far from most of the shale gas well sites. Therefore, field measurements are highly desired, and investigation of actual ambient concentrations of $PM_{2.5}$ around these sites would be the subject of a valuable future study.

Conclusion

Results from this research indicate that current Pennsylvania setback policy of 152.4 m (500 ft) is inadequate to protect residents from exceedances of the EPA's daily concentration standard for $PM_{2.5}$, and it is inadequate to protect against exceedances of the annual concentration standard for sites with 6 or more wells. To protect occupied buildings and outdoor areas against exceedances of the daily average standard, this analysis suggests that setback distances need to be up to 736 m. To protect against exceedances of the annual average $PM_{2.5}$ standard, setback distances should be a function of the number of wells drilled at the site. Further refinements to this analysis are needed to account for multiple pollutants. Alternative policy options include limits on the number of wells per site (a choice that may have negative environmental implications, as it would increase the number of constructed well pads) and limiting the maximum $PM_{2.5}$ emission rate at each site to 0.165 kg/hr.

The results provided here are associated with a generic well site in Pennsylvania with no specific structure within its vicinity. If there is such a structure close to a well site of interest or for an unusual case like an inhabited building, it would be more prudent not to generalize these results to that case. For such a case, there would be a need for a specific analysis of the site and its vicinity. In addition to buildings, large trees and geographic terrain (hills, slopes, cliffs, ravines, canyons, etc.) could also have specific effects on downwind concentrations not addressed here. The effects of these obstructions are diverse, raising and lowering concentrations outside the setback area in a manner that is not readily generalizable for mitigating policy. We recommend that for locations that diverge significantly from the assumptions in this analysis, a specific analysis should be conducted to establish the appropriate setback distance. Also, it is worth repeating that the emission rate corresponding to the conservative case in this work is the 97.5th percentile. Simulation of the most conservative case requires the necessary update of this value.

Acknowledgment

The authors are grateful to Dr. Sanjay Srinivasan and Dr. John Y. Wang for their insights into unconventional shale gas development processes and procedures.

Funding

This work was supported by the John and Willie Leone Family Department of Energy and Mineral Engineering, College of Earth and Mineral Sciences, Pennsylvania State University.

About the authors

Zoya Banan is an is a Ph.D. candidate in energy management and policy at the Department of Energy and Mineral Engineering, Pennsylvania State University, University Park, PA.

Jeremy M. Gernand is an assistant professor of industrial health and safety at the Department of Energy and Mineral Engineering, Pennsylvania State University, University Park, PA.

References

Adgate, J.L., B.D. Goldstein, and L.M. McKenzie. 2014. Potential public health hazards, exposures and health effects from unconventional natural gas development. *Environ. Sci. Technol.* 48:8307–8320. doi:10.1021/es404621d.

Akpootu, D.O., and N.N. Gana. 2013. The effect of relative humidity on the hygroscopic growth factor and bulk hygroscopicity of water soluble aerosols. *Int. J. Eng. Sci.* 2:48–57.

Arystanbekova, N.K. 2004. Application of Gaussian plume models for air pollution simulation at instantaneous emissions. *Math. Comput. Simul* 67:451–58. doi:10.1016/j.matcom.2004.06.023.

Boxall, P.C., W.H. Chan, and M.L. McMillan. 2005. The impact of oil and natural gas facilities on rural residential property values: A spatial hedonic analysis. *Resour. Energy Econ.* 27:248–269. doi:10.1016/j.reseneeco.2004.11.003.

Compilation of Air Pollutant Emissions Factors. 1972. http://www.epa.gov/ttn/chief/ap42/index.html (accessed November 22, 2016).

Connolly, P. 2014. http://personalpages.manchester.ac.uk/staff/paul.connolly/teaching/practicals/gaussian_plume_modelling.html. (accessed October 10, 2016).

Corbett, T., and E.C. Abruzzo. 2014. *Commonwealth of Pennsylvania Department of Environmental Protection 2014 Annual Ambient Air Monitoring Network Plan*. Harrisburg, Pennsylvania: Pennsylvania Department of Environmental Protection.

Department of Oil and Gas Reporting website. 2016. http://www.depreportingservices.state.pa.us/ReportServer/Pages/ReportViewer.aspx?/Oil_Gas/Permits_Issued_Detail. (accessed December 24, 2016).

EIA—Shale gas production. 2016. http://www.eia.gov/dnav/ng/ng_prod_shalegas_s1_a.htm (accessed November 22, 2016).

EIA—Shale gas proved reserves. 2016. http://www.eia.gov/dnav/ng/ng_enr_shalegas_a_epg0_r5301_bcf_a.htm (accessed November 22, 2016).

EPA. 2014 National emissions inventory (NEI). https://www.epa.gov/air-emissions-inventories/2014-national-emissions-inventory-nei-data (accessed March 10, 2017).

EPA. 2016. *Office of air quality planning and standards, air quality assessment division, air quality modeling group*. User's Guide for the AERMOD Terrain Preprocessor (AERMAP). Research Triangle Park, North Carolina: U.S. Environmental Protection Agency.

EPA. 2017a. Outdoor air quality data—monitored values report. Annual Data. https://www.epa.gov/outdoor-air-quality-data/monitor-values-report (accessed December 19, 2017).

EPA. 2017b. Outdoor air quality data—daily data. https://www.epa.gov/outdoor-air-quality-data/download-daily-data (accessed December 19, 2017).

Fry, M. 2013. Urban gas drilling and distance ordinances in the Texas Barnett Shale. *Energy Policy* 62:79–89. doi:10.1016/j.enpol.2013.07.107.

Gopoch, A., S. Burk, and K.L. Davidson. 1980. Stability effects on aerosol size and height distributions. *Tellus* 32 (3):245–250.

Haley, M., M. McCawley, A.C. Epstein, B. Arrington, and E.F. Bjerke. 2016. Adequacy of current state setbacks for directional high-volume hydraulic fracturing in the Marcellus, Barnett, and Niobrara shale plays. *Environ. Health Perspect.* 124 (9):1323–1333. doi:10.1289/ehp.1510547.

IEM: Iowa Environmental Mesonet. 2016. http://mesonet.agron.iastate.edu/request/download.phtml?network=PA_ASOS (accessed November 22, 2016).

Jacoby, H.D., F.M. O'Sullivan, and S. Paltsev. 2011. The influence of shale gas on U.S. Energy and Environmental Policy. *Econ. Energy Environ. Policy* 1 (1):37–51.

Jenner, S., and A.J. Lamadrid. 2013. Shale gas vs. coal: Policy implications from environmental impact comparisons of shale gas, conventional gas, and coal on air, water, and land in the United States. *Energy Policy* 53:442–453. doi:10.1016/j.enpol.2012.11.010.

Kelly, F.J., and J.C. Fussell. 2012. Size, source and chemical composition as determinants of toxicity attributable to ambient particulate matter. *Atmos. Environ.* 60:504–526. doi:10.1016/j.atmosenv.2012.06.039.

Lepeule, J., F. Laden, D. Dockery, and J. Schwartz. 2012. Chronic exposure to fine particles and mortality: An extended follow-up of the Harvard six cities study from 1974 to 2009. *Environ. Health Perspect.* 120 (7):965–970. doi:10.1289/ehp.1104660.

Martin, A.R., and W.H. Finlay. 2005. The effect of humidity on the size of particles delivered from metered-dose inhalers. *Aerosol Sci. Technol.* 39 (4):283–289. doi:10.1080/027868290929314.

National Ambient Air Quality Standards (NAAQS). 2012. 40 CFR part 50. Washington, DC.

Ogoke, V., L. Schauerte, G. Bouchard, and S.C. Inglehart. 2014. Simultaneous operations in multi-well pad: A cost effective way of drilling multi wells pad and deliver 8 fracs a day. Proceeding of the Society of Petroleum Engineers Annual Technical Conference and Exhibition, Amsterdam, The Netherlands, October 27–29.

Olaguer, E.P. 2012. The potential near-source ozone impacts of upstream oil and gas industry emissions. *J. Air Waste Manage. Assoc.* 62 (8):966–977. doi:10.1080/10962247.2012.688923.

Pacsi, A.P., N.S. Alhajeri, D. Zavala-Araiza, M.D. Webster, and D.T. Allen. 2013. Regional air quality impacts of increased natural gas production and use in Texas. *Environ. Sci. Technol.* 47:3521–3527. doi:10.1021/es3044714.

Peterson, R.L., S.A. Guerra, and A.S. Bova. 2017. Critical review of the building downwash algorithms in AERMOD. *J. Air Waste Manage. Assoc.* 67 (8):826–835. doi:10.1080/10962247.2017.1279088.

Peterson, R.L., and A. Beyer-Lout. 2012. Aermod building downwash theoretical limitations and possible solutions. Paper presented at 105th Annual Conference and Exhibition of the Air & Waste Management Association, San Antonio, Texas, June.

Petters, M.D., and S.M. Kreidenweis. 2007. A single parameter representation of hygroscopic growth and cloud condensation nucleus activity. *Atmos. Chem. Phys.* 7:1961–1971. doi:10.5194/acp-7-1961-2007.

Popovicheva, O.B., N.M. Persiantseva, N.K. Shonija, P. DeMott, K. Koehler, M. Petters, S. Kreidenweis, B. Demirdjian, J. Suzanne, and V. Tishkova. 2008b. Water interaction with hydrophobic and hydrophilic soot particles. *PCCP* 10:2332–2344. doi:10.1039/b718944n.

Popovicheva, O.B., N.M. Persiantseva, V. Tishkova, N.K. Shonija, and N.A. Zubareva. 2008a. Quantification of water uptake by soot particles. *Environ. Res. Lett.* 3:025009. doi:10.1088/1748-9326/3/2/025009.

Rodriguez, M.A., M.G. Barna, and T. Moore. 2009. Regional impacts of oil and gas development on ozone formation in the Western United States. *J. Air Waste Manage. Assoc.* 59 (9):1111–1118. doi:10.3155/1047-3289.59.9.1111.

Roohani, Y.H., A.A. Roy, J. Heo, A.L. Robinson, and P.J. Adams. 2017. Impacts of natural gas development in the Marcellus and Utica shales on regional ozone and fine particulate matter levels. *Atmos. Environ.* 155:11–20. doi:10.1016/j.atmosenv.2017.01.001.

Roy, A.A., P.J. Adams, and A.L. Robinson. 2014. Air pollutant emissions from the development, production, and processing of Marcellus Shale natural gas. *J. Air Waste Manage. Assoc.* 64 (1):19–37. doi:10.1080/10962247.2013.826151.

Shah, S.D., D.R. Cocker, J.W. Miller, and J.M. Norbeck. 2004. Emission rates of particulate matter and elemental and organic carbon from in-use diesel engines. *Environ. Sci. Technol.* 38 (9):2544–2550. doi:10.1021/es0350583.

Shonkoff, S., J. Hays, and M.L. Finkel. 2014. Environmental public health dimensions of shale and tight gas development. *Environ. Health Perspect.* 122 (8):787–795.

Sinclair, D., R.J. Countess, and G.S. Hoopes. 1974. Effect of relative humidity on the size of atmospheric aerosol particles. *Atmos. Environ.* 8:1111–1117. doi:10.1016/0004-6981(74)90045-6.

Touma, J.S., V. Isakov, J. Ching, and C. Seigneur. 2006. Air quality modeling of hazardous pollutants: Current status and future directions. *J. Air Waste Manage.* 56 (5):547–558. doi:10.1080/10473289.2006.10464480.

Vidic, R.D., S.L. Brantley, J.M. Vandenbossche, D. Yoxtheimer, and J.D. Abad. 2013. Impact of shale gas development on regional water quality. *Sci.* 340 (6134):1235009. doi:10.1126/science.1235009.

Winkler, P. 1988. The growth of atmospheric aerosol particles with relative humidity. *Physical Scr.* 37:223–230. doi:10.1088/0031-8949/37/2/008.

World Health Organization (WHO). 2003. *Health aspects of air pollution with particulate matter, ozone and nitrogen dioxide*. Bonn, Germany: The WHO Regional Office for Europe.

Zielinska, B., E. Fujita, and D. Campbell. 2011. *Monitoring of emissions from barnett shale natural gas production facilities for population exposure assessment*. Houston, TX: Desert Research Institute.

ATTACHMENT C

STUDY 19

PREPRINT

Emissions of Particulate Matter due to Marcellus Shale Gas Development in Pennsylvania: Mapping the Implications

Zoya Banan [a], Jeremy M. Gernand [a,]*

[a] Department of Energy and Mineral Engineering, Pennsylvania State University, University Park, PA 16802, USA.

* Corresponding author. Tel.: +1(814)865-5861.

E-mail addresses: zfb5010@psu.edu (Z. Banan),
jmgernand@psu.edu (J.M. Gernand).

Abstract

Over the past decade, the shale gas boom has led to increasing public concerns regarding the effects of population exposure to air pollutants from shale gas development activities with concentrations higher than the EPA's National Ambient Air Quality Standards originating. This study investigates the sufficiency of current policy in Pennsylvania in protecting people from exposure to levels of fine particulate matter ($PM_{2.5}$) which exceed this standard. We used a Gaussian plume model to simulate the fine particulate matter ($PM_{2.5}$) concentrations over the Marcellus shale region of Pennsylvania, considering the temporal and spatial density of these activities between 2005 and 2017. Simulation results were synced with census block data to estimate the potential number of people who experienced exceedances of the $PM_{2.5}$ air quality standards during this period. Results demonstrate that setback policy in Pennsylvania may not be adequate to maintain the exposure level in residential areas below the standard. Emissions from shale gas development alone (not accounting for background concentration) could cause up to 174 persons in one year (2015) to experience concentrations higher than the EPA's annual standard for $PM_{2.5}$. The additional number of exceedances from shale gas development to those attributable to background is estimated to raise up to more than 36,000 persons in a single year which is almost 1% of the Marcellus Shale regional population in Pennsylvania. Findings indicate that the number of affected residents has largely been proportional to the overall number of developed wells in the state, but specific development histories in some counties and in some years show how similar levels of development could occur with reduced population exposure. Setback distance is shown to be an effective method to reduce some exposure exceedances, but it should be revised based on the number of wells per wellpad as well as the local conditions to further limit air quality impacts.

Keywords: Setback policy, Community air quality, Emission exposure, Gaussian plume model

1. Introduction

Over the past two decades, the technological innovations in horizontal and directional drilling and hydraulic fracturing led to a shale gas boom as they made development of shale reserves more economical. Shale plays are shale formations where significant amount of natural gas has accumulated. In North America, the Marcellus Shale play is known as one of the most prolific plays in the United States. Data show that production from Marcellus shale gas in Pennsylvania increased from 1.1 trillion cubic feet in 2011 to more than 5 trillion cubic feet in 2016 (EIA – Shale gas production, 2018).

From 2005 to 2017, more than 18,000 shale gas wells were permitted for development in the Marcellus Shale region of Pennsylvania (Department of Oil and Gas Reporting website, 2018). The increase in shale gas production has moved the shale gas wells closer to residential areas (Adgate et al., 2014). This is not true only for the Marcellus Shale, but other regions as well. For instance, McKenzie et al. (2016) stated that the number of people who lived less than 500 ft. from an active oil and gas well in Denver-Julesburg Basin, Colorado approximately increased from 3,800 to 8,900 persons between the years 2000 and 2012.

Large diesel-powered equipment and gas turbines are used during the drilling and hydraulic fracturing stages of shale gas development that emit direct and fugitive air emissions (ICF International, 2009; McKenzie et al., 2012). Studies have shown that these emissions can affect the air quality within local areas around shale gas development sites and even farther downwind (Cheng et al., 2015; Gilman et al., 2013; Moore et al., 2014) which has been associated with an increased risk of health issues (Adgate et al., 2014; Croft et al., 2019; Krzyzanowski, 2012). The identified emissions from shale gas activities include but not limited to volatile organic compounds (VOCs), nitrogen dioxide, sulfur dioxide, and particulate matter (PM) (ICF International, 2009;

Zielinska et al., 2011). The World Health Organization has announced serious concern regarding the human health burden of fine particulate matters ($PM_{2.5}$) (WHO, 2007) and the finding of epidemiologic studies indicated an association between exposure to $PM_{2.5}$ emissions and increase in the incidences of acute health outcomes, such as asthma (Rasmussen et al., 2016), cardiological and neurological diseases (Jemielita et al., 2015) and upper respiratory health impacts (Rabinowitz et al., 2015). Nevertheless, very few studies have investigated the local residents' exposure to PM emissions from shale gas developments.

Federal and state governments have tried to control the level of human exposure to dangerous pollution levels through consideration of some regulations. The United States Environmental Protection Agency (EPA) tried to put control on exposure to six pollutants known as "criteria" pollutants through National Ambient Air Quality Standards (NAAQS). These pollutants are nitrogen dioxide (NO_2), sulfur dioxide (SO_2), particulate matter (PM), ozone (O_3), lead (Pb), and carbon monoxide (CO). However, these standards fail to address short-term exposure to high pollutant concentrations, because they are designed based on concentration averages over the long-term and aimed at determining the compliance of major pollution sources with permitted emissions (Brown et al., 2014).

Consideration of setback distances is another effort to provide human health protection at the state and local level. Setback distance indicates the minimum distance that any occupied residential buildings should be located from a shale gas wellsite. Setback distances are known to be primarily decided based on political compromises among governments, landowners, and environmental groups (Haley et al., 2016). Haley et al. (2016) states that the current setback requirement in the Marcellus Shale of Pennsylvania is not sufficient to maintain human exposure below the established limits for benzene and hydrogen sulfide. According to Banan and Gernand

(2018), the existing setback policy of 152.4 m (500 ft.) in Pennsylvania (58 PA Cons Stat § 3215) is not effective in every direction from a typical wellpad (6 wells per pad) to protect the residential areas from $PM_{2.5}$ concentrations above the EPA NAAQS.

The distance of a shale gas well from residential areas is a critical factor in evaluation of the risk to human health caused by hydraulic fracturing activities (Meng and Ashby, 2014). The distance-based risk analysis by Meng (2015) found 3% of the shale gas wells to impose a high level of risk to the population who live within 1 km of a shale gas well. Other studies have shown that residents living within 0.8 km of a shale gas well face greater health risks than the ones living farther away (McKenzie et al., 2012; Meng and Ashby, 2014). However, the importance of distance from shale gas development wellsites with respect to air quality changes has been the scope of a few recent studies. Exposure intensity, frequency and duration are critical factors in exposure evaluation as well. It is also important to consider the aggregated impact of emissions from multiple sources placed near a residential area (Brown et al., 2015). These factors are some of the major uncertainties regarding the public health effects of shale gas operations (Adgate et al., 2014) and to the best of our knowledge, there is no study which addresses these uncertainties.

The goal of this study is to investigate the suitability of current Pennsylvania setback policy for shale gas development with respect to protecting people from exposure to $PM_{2.5}$ concentrations higher than NAAQS. This analysis focuses on $PM_{2.5}$ emissions as a less investigated air pollutant by previous studies in association with shale gas developments and accounts for the intensity, frequency and duration of residents' exposure to these emissions in evaluation of setback policy. We simulate $PM_{2.5}$ emissions dispersion from development of the wells in Marcellus Shale gas in Pennsylvania between 2005 and 2017 and evaluate the associated influence on local air quality. The latest census data (Census Data, 2010) synced with the simulated concentrations allows estimation of the potential number of people who experienced concentration levels higher than the EPA's annual standard. Spatial and temporal simulation of shale gas emissions dispersion in this study helps to shed light on the role of density of well development and well locations on the community's exposure exceedance within the vicinity of wells and how such exposures have changed over time.

2. Methodology

2.1. Data Sources

The first step of this study is to simulate $PM_{2.5}$ concentrations within the vicinity of each of the developed wells from 2005 to 2017 in the Marcellus shale region of Pennsylvania. We then compare the annual average concentrations with EPA's latest standard, an annual mean $PM_{2.5}$ concentration of 12 $\mu g/m^3$ (NAAQS, 2012). This number is the primary annual standard which was considered by EPA to provide public health protection. EPA's NAAQS also indicate a cap of 35 $\mu g/m^3$ on the daily mean $PM_{2.5}$ concentrations (NAAQS, 2012). However, this study focuses on the annual mean concentration of these emissions with respect to shale gas developments which would be informative for later evaluation of health risks and outcomes in association with exposure to emissions from such sources.

The source for shale gas well data are the reports of issued permits by the Pennsylvania Department of Environmental Protection (Department of Oil and Gas Reporting website, 2018). According to this dataset, shale gas developments occurred in 39 counties in Pennsylvania between 2005 and 2017. The well data from this source relevant to the purposes of this study are location, number of wells per wellpad, and "SPUD Date". According to EPA Oil and Gas Dictionary (2018), "SPUD Date" is defined as the date that drilling of the well has commenced. We also extracted the data on drilled depth ("TotalDepth") and depth of fracking ("UpperPerf") from Drillinginfo (2019) for the wells that were permitted and developed between 2005 and 2017 in Pennsylvania. To calculate the length of fracturing for each drilled well, we subtracted the depth of fracturing from the drilled depth. Data from DrillingInfo was used to normalize the period of drilling and fracking for each well based on daily rates of drilling and fracking. “SPUD Date” is not available for all permitted shale gas wells. We assumed that development never started at these wells.

We used measured wind data at the closest monitoring station to simulate emission dispersion from each wellsite during the period of development. Wind data (including wind direction, wind speed, relative humidity, and cloud coverage) were available for almost every hour from 2005 to 2017 at seventeen weather monitoring stations in Pennsylvania (IEM, 2018). For missing hourly data, we used bootstrapping of the available data from the week before and from the same calendar date and hour of other years for the same station to estimate the values. Wind roses for these stations are available in Supplemental File 2 of this manuscript.

Due to the lack of direct measurements of pollutant emission rates from shale gas activities, we used estimated $PM_{2.5}$ emission rates by Roy et al. (2014). They used the reported emission factors from EPA’s inventory models (AP-42) and values in the literature reported for diesel engines similar to those that are being used in drilling and fracturing of Marcellus shale gas wells. They quantified the variables of emission equations (e.g., emission factors, engines’ horsepower, load factor, number of fracking stages, etc.) by means of a Monte Carlo simulation considering the specific distribution of each variable. We recalculated the reported emission rates by Roy et al. (2014) for the drilling step in grams per hour and for the fracking step in grams per stage (instead of tons per year per well drilled). We omitted the terms for “time to drill one well” and “number of stages” from the equations pertaining to drilling and fracturing steps, respectively, in the Monte Carlo simulations. Therefore, results from our Monte Carlo simulations were the hourly rates of drilling and hydraulic fracturing. We used these hourly emission rates in combination with specific operation duration for each well with respect to its drilled depth and length of fracking. Since detailed information about activities timing is not publicly available, we estimated a drilling rate of 1000 feet per day based on the average drilled depth (12,546 ft., from well dataset) and drilling time (Facts about Canada's Oil and Natural Gas Industry, 2019). It is estimated that it takes one day in average to fracture 1000 feet in three stages (McKeon, 2011; Facts about Canada's Oil and Natural Gas Industry, 2019; Coloradans for Responsible Energy Development, 2019). We estimated the mean and 95% CI for $PM_{2.5}$ emission rate from drilling and hydraulic fracturing to be 0.45 (0.1 – 1.3) and 1.06 (0.28 – 1.88) kilogram per hour, respectively. In this article, we discuss the estimated exposure exceedances in association with dispersion $PM_{2.5}$ emissions from shale gas developments at high hourly emission rate values (97.5th percentile) and provide an upper estimation of such exceedances in this regard. More details on estimated number of exposure exceedances at mean and low (2.5th percentile) levels of hourly emission rates over the period of study are available in the supplemental file to this manuscript (Table-4S).

We used U.S. Census block-level population data to estimate the number of people who might be exposed to concentration levels higher than NAAQS for $PM_{2.5}$. The latest U.S. Census block data (Census Data, 2010) including population, block area, block geographic location (latitude and longitude) was synced with the simulated concentrations and their locations.

There are other types of stationary and mobile sources of $PM_{2.5}$, such as motor vehicles, factories, and roads. Temporary or permanent dispersion of emissions from these kind of sources leads to various levels of background concentrations at different locations. EPA provides summary reports of daily $PM_{2.5}$ concentrations measured at individual monitoring sites in Pennsylvania (EPA, 2018). We considered the reported measurements by EPA for the background concentration in our evaluations in order to identify the locations where concentration level passed the limit due to nearby shale gas activities.

2.2. Model

For the purpose of this research, we used a Gaussian plume model as the simulation method to estimate PM concentrations at the local scale. The Gaussian plume model is an advection-diffusion equation that describes the movement of a plume in the atmosphere mainly within 50 km from the source. The model is formulated in equation (1):

$$C(x,y,z) = \frac{Q}{2\pi u \sigma_y \sigma_z} \, exp\left(-\frac{y^2}{2\sigma_y^2}\right)\left[exp\left(-\frac{(z-H_{eff})^2}{2\sigma_z^2}\right) + \, exp\left(-\frac{(z+H_{eff})^2}{2\sigma_z^2}\right)\right] \quad (1)$$

where C is the substance concentration at any point of x, y, and z, where x is the distance downwind from the well, y is the crosswind distance from the plume centerline, and z is the vertical distance from the ground level. Q is emission rate of the substance from the source, u is the average wind speed at wellsite height, H_{eff} is the effective wellsite height, σ_y is dispersion coefficient in the crosswind direction, and σ_z is dispersion coefficient in the vertical direction. The term in the brackets rules the distribution of plume in vertical dimension (z) at a given downwind distance (x) considering the effect of surface reflection. This implies that Gaussian plume model assumes no dry and wet deposition.

The sources of $PM_{2.5}$ emissions from shale gas activities are mainly diesel engines (ICF International, 2009; Roy et al., 2014). These engines are usually located at the ground level and they emit exhaust at different heights and speed, but detailed information is not publicly available on these factors. In this study, we assumed that H_{eff} is the same value for all simulation cases and that the model simulates concentrations at the vertical height (z) equal to H_{eff}. Accounting for the wellpad elevation (typically 4 to 5 feet) as well as 1 to 2 feet elevation of the plume center, we assumed that the engines emitted emission plume centered at the height of 2 meters.

The model takes atmospheric stability as input to calculate parameters σ_y and σ_z at different distances downwind from the source (EPA, 1995). We defined the daytime (night-time) between 5 am to 8 pm (between 8 pm to 5 am) for each 24-hour period and identified stability categories using the Turner method based on wind speed, cloud coverage data, and time of the day (Turner, 1970).

We used MATLAB to implement the Gaussian plume model that simulates PM concentrations at every hour of operation for each developed wellsite. Our model is developed based on the model code developed by Banan and Gernand (2018). However, it is modified to simulate the emission dispersion from developed shale gas wellsites, instead of providing a probabilistic evaluation of exceedance cases for a representative one. Our model accounts for the role of time, emission characterizations, relative humidity, and atmospheric stability.

Our model simulates $PM_{2.5}$ emissions from diesel engines located at the site and does not account for fugitive dust, mineral dust from proppant handling, or emissions from trucks and other vehicles. We aim to model $PM_{2.5}$ concentration at distances greater than 152.4 m (500 ft.) from the well location and therefore, it is reasonable to assume that all emissions are originated from the center of the activity. Presence of buildings and other structures around the emission source could affect emission dispersion in the near field (Petersen et al., 2017). However, shale gas development activities in Pennsylvania generally occur in rural areas and as a result, the probability of existence of such structures within very close vicinity of the wellsites is low. Moreover, studies which took advantage of simulating software, such as AERMOD, with ability to account for the changes in the terrain, reported the simulation results to include both overprediction and underprediction of actual concentrations under different input setups, compared to the case of flat terrain (Amoatey et al., 2019; Matacchiera et al., 2019; Tartakovsky et al., 2013). Tartakovsky et al. (2013) stated the better performance of AERMOD over CALPUF in predicting the level of total suspended particulates dispersed from quarries to be coincidental and very volatile with respect to the input meteorological dataset. Thus, we assumed a flat terrain in this evaluation.

2.3. Analysis

In this study, we simulated the dispersion of $PM_{2.5}$ emissions from development activities at every shale gas wellsite at each hour of operation during drilling and hydraulic fracturing stages, using a Gaussian plume model. For the purpose of this study, we accounted for the $PM_{2.5}$ emissions from the typical diesel-powered engines which are being used at these sites; i.e., pumps powered by off-road heavy-duty diesel engines, diesel engine generators, and diesel-powered compression ignition engines (ICF International, 2009; Roy et al., 2014). The model calculates the average $PM_{2.5}$ concentration over a year. The simulation results demonstrate the locations where annual mean concentration of $PM_{2.5}$ exceeded the EPA's annual standard for $PM_{2.5}$. Then, we estimated the number of people who lived in those locations using the United States Census block data from 2010.

The output of this study sheds light on discussions around the efficiency and sufficiency of current setback policy in Pennsylvania. For the purpose of this study, we assumed that Pennsylvania setback policy was rigorously enforced in the Marcellus shale gas region and no people were living closer than 154 m (500 ft.) to the shale gas wellsites during their development period. We used the annual average concentrations at each location to estimate the potential number of people who experienced concentrations above the standard.

For each developed wellsite, the activities time window is in accordance with the total drilled depth of wells per wellpad. We used a grid size of 25-by-25 meters to specify the locations around each wellsite where the model simulates the concentrations. The model includes the wind data measured during the identified time window at the closest monitoring station to each wellsite and generates arrays of $PM_{2.5}$ concentrations for all grids on an hourly basis. Our model accounts for cumulative aspect of emission dispersion due to simultaneous development activities in one area at each hour. Thus, the final matrix of concentrations contains the cumulative concentrations generated from all wellsites per each hour at every single location. The third dimension represents all hours of one year. To identify the locations where concentrations exceed the annual standard, the model calculates the mean concentrations over all the hours of one year (i.e., the third dimension of the concentration matrix). In this case, exceedance occurs where the annual average concentration exceeds the primary annual standard, which is 12 $\mu g/m^3$ for $PM_{2.5}$. We assumed that population was evenly distributed in each census block. The model multiplies the population density of the closest block (persons per area of the block) to each grid by the grid size to provide an estimation of its population.

According to EPA NAAQS, compliance with $PM_{2.5}$ annual standard is determined based on the average of annual mean concentrations over three consecutive years. However, the temporary nature of shale gas development activities does not allow for evaluation of $PM_{2.5}$ concentrations over a three-year period. As this work aims to evaluate the air quality status from public health aspect, we consider the averaging time periods of one year in order to investigate any exceedance of the annual standard.

Our model simulates the emission dispersion from development of each wellsite up to 8 km downwind of the source. According to Touma et al. (2006), it is appropriate to apply a Gaussian plume model over this distance, as it can be used for dispersion modeling within 50 km from the source. Also, it accounts for the influence of relative humidity on the size of particles. Relative humidity causes an increase in the particle size in accordance to its properties, such as particle type, dry particle size, and level of humidity (Gopoch et al., 1980; Martin and Finlay, 2005; Sinclair et al., 1974; Winkler, 1988).

This model adds the relevant background concentrations to the modeled local concentration at locations where the air quality is affected by shale gas development activities. It uses the measurements at the EPA's air quality monitoring stations to each grid. These monitoring stations are not located within the vicinity of most of the developed shale gas wellsites. Overall, they were within 0.7 to 140 km from the developed wells over the time window of this study (mean, median, and mode values were 49, 38 and 15 km, respectively). Results from cross-correlation of ambient $PM_{2.5}$ measurements for any pair of these stations demonstrate a median correlation coefficient of 0.74. Also, 10% of the pairs have correlation coefficient of less than 0.5 and they pertain to the cases with limited available measurements at either one or both of the paired stations or to the case of pairs located farther away from each other in the state. Nevertheless, the measurements at these stations may not precisely represent the background concentrations in development areas and might introduce more uncertainty due to underestimation or overestimation of background level at different locations. Background concentration plays a significant role in identification of the areas where exceedance of NAAQS may occur. Thus, we took advantage of the inverse distance weighting (IDW) method to calculate the average annual background concentration using the measurements at the three nearest EPA's air quality monitoring stations to account for its significant impact.

Well data contains few cases where "SPUD Date" is reported, but data on total drilled depth and/or depth of fracking is missing. The model conducts a bootstrapping simulation on the available data, from drilling of other developed wells in the same county in the same year, to simulate the drilled depth for these wells. We used the average drilled depth from all drilled Marcellus shale gas wells in Pennsylvania (12000 feet) for one wellsite in Clarion County (in 2017), as no data were available from nearby wells. Refer to Figure-1S for more details on average drilled depth. To simulate the missing data on the length of fracking, we conducted a simple linear regression between total drilled depth and length of fracking. The regression resulted in equation (2), as follows:

$$\text{Fracking Length} = 0.7313 \times \text{Total Drilled Depth} - 4371.3 \quad (2)$$

where Fracking Length and Total Drilled Depth are in the unit of feet (see Figure-2S for more details). Even though longer drilled depth may not necessarily cause a longer length of fracking, data from the current practicing in Marcellus shale region demonstrate such a tendency (p-values are both less than 0.00001 for the intercept and the coefficient). Therefore, we used this equation to estimate the missing values for length of fracking. Refer to supplemental file, Table-7S for the ANOVA table. For the wells with total drilled depth less than 6000 feet, length of fracking was determined using bootstrapping of the corresponding values from wells shorter than 6000 feet deep (63 cases).

The model accounts for four different metrics to estimate the size of the affected population: 1) exceedance occurring due to shale gas development only (zero background concentration), 2) exceedance occurring due to shale gas emissions plus background concentration, 3) exceedances occurring due to the background concentration alone (zero shale emission), and 4) the additional exceedances occurring due to shale gas emissions from background only (i.e., metric

3). In other words, metric 2 is the summation of metric 3 and metric 4. As an example, if the exceedances occurring due to the background concentration alone (metric 3) in a specific location was 100 persons and this number increased to 120 persons when shale gas emissions were added to the background level (metric 2), the additional exceedances occurring due to shale gas emissions from background only (metric 4) would be 20 persons. These categories help to clarify the number of people who became exposed to PM concentrations higher than EPA's standard due to shale gas activities.

Our model assumes the same hourly emission rates (discussed previously) for all wells during drilling and hydraulic fracturing stages. However, the emission rate may vary depending on engines specifications, technology and operational remedies (such as changing engines' fuel to natural gas at completion of first well) used at each wellsite, and in that regard, the estimated values used in this study could be an under- or over-estimation of the true rate for any well. Moreover, while operation time is also a function of weather delays, this information is not available in the well dataset, and the actual duration could be longer or shorter than the estimated value. Also, along with typical assumptions in Gaussian plume dispersion modeling, this model does not consider the effects of rain or wet ground surfaces.

Weather monitoring stations were not located close to all the developed shale gas wells; the closest one was 0.5 km away from a developed well. Data indicates that this distance was 96.5 km in the farthest case (mean and median distances were 40.3 km and 39.6 km, respectively). Wind profiles might not remain the same over such a distance, but wind data measurements are not available at the location of every developed well. This study assumes constant wind speed and direction over the concentration simulation area during each hour. Also, the model assumes that there is no accumulation of emissions from hour to hour.

3. Results

Figure-1 depicts the annual average concentration of $PM_{2.5}$ emissions simulated for the developed wells in Washington County, PA between the years 2006 and 2017. The color coding helps to identify the locations where the average total concentration was higher than the annual EPA standard for $PM_{2.5}$ (12 $\mu g/m^3$). The maps capture the density and movement of development activities in this county over time. The term "density" includes the number of wells per each wellpad (wellpad density) as well as the number of wells being developed per unit area (well area-density).

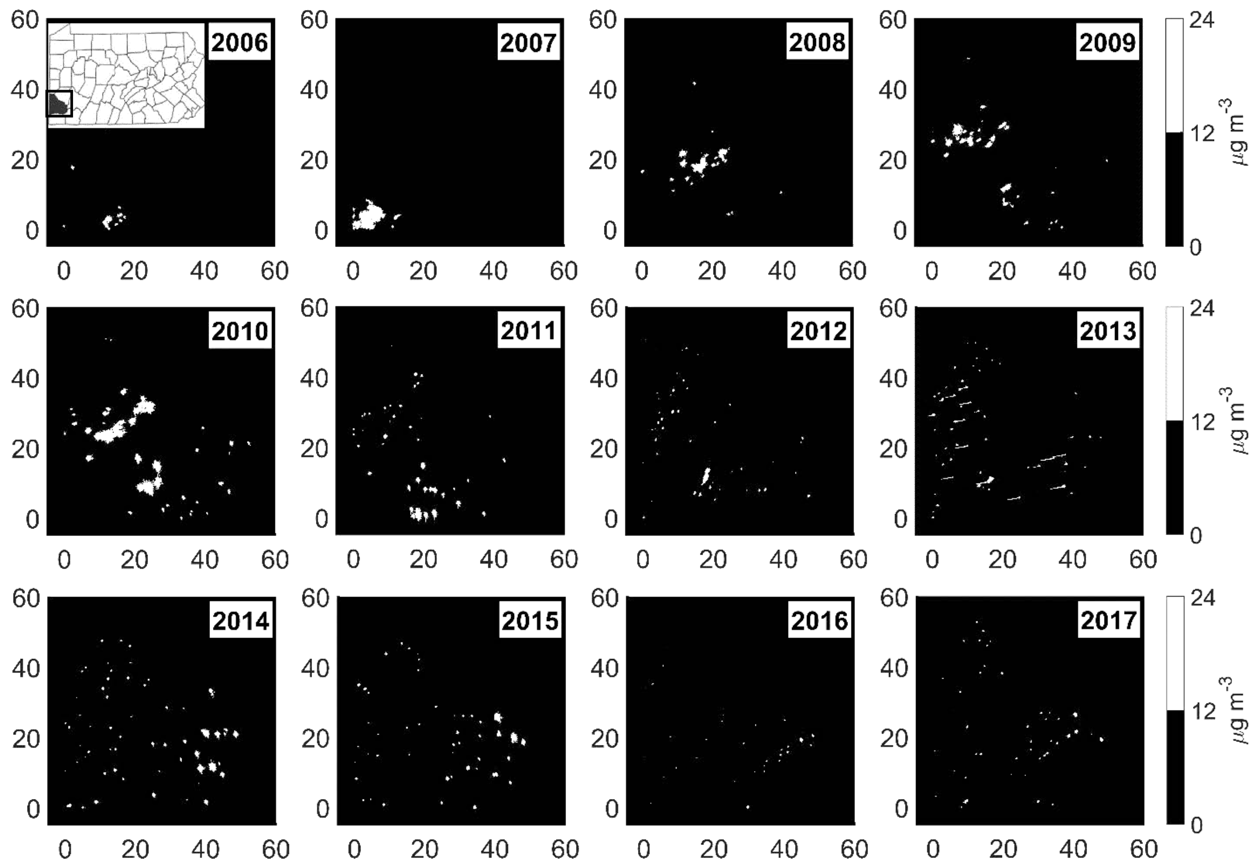


Figure-1: Map of $PM_{2.5}$ emissions from shale gas wells developed during 2006 to 2017 in Washington County, PA. Map depicts the view from top and scale is in kilometers. The location of Washington County is shown on the Pennsylvania map at the upper left corner. The black square on the map shows the area of simulation in different years.

According to well permit records, similar numbers of wells were developed in Washington County, in 2010 and 2011 (166 and 155, respectively). Figure-1 demonstrates greater areas with concentrations higher than standard in 2010 than in 2011. Defining high density to be 6 or more wells per 1 km^2 and considering only nonzero well area-density grids, histograms of well area-density for Washington County showed higher frequency of high well area-density in 2010 than in 2011 (28% versus 18%, respectively) (refer to Figure-3S for histograms).

Shale gas activities in Washington County could have caused 8,156 persons to become exposed to concentrations higher than the standard in 2010. This number is estimated to be only 1,555 people in 2011. This suggests that there were cases wherein similar numbers of wells in more densely developed wellpads affected less population by poor air quality (81% reduction in this case). Evaluation of population density within the vicinity of developed wells in these two years indicates that shale gas developments in Washington County in 2010 were closer to more populated areas. The median value of population for nearby blocks in this county in 2010 and 2011 were 28 and 27 persons/km^2, respectively. However, the maximum values were 1,386 and 205 persons/km^2, respectively. Moreover, the median and mean wellpad densities in 2010 were 2 and 3.1 wells per wellpad, respectively (refer to Table-3S). These values were equal to 4 and 3.9 wells per wellpad in 2011.

Proximity to residential areas plays a significant role in changing the number of people experiencing exceedances and would make the development of a wellsite with lower wellpad density cause more negative impacts on the local air quality than the one with higher wellpad density. Figure-2 depicts two wellsites with different wellpad densities located within the vicinity of areas with different population densities. This figure depicts the average $PM_{2.5}$ concentrations over one year for each case, using the hourly levels simulated by our Gaussian plume model. The color-code helps to indicate the locations within the vicinity of each demonstrated wellsite where annual average concentration exceeded 12 $\mu g/m^3$.

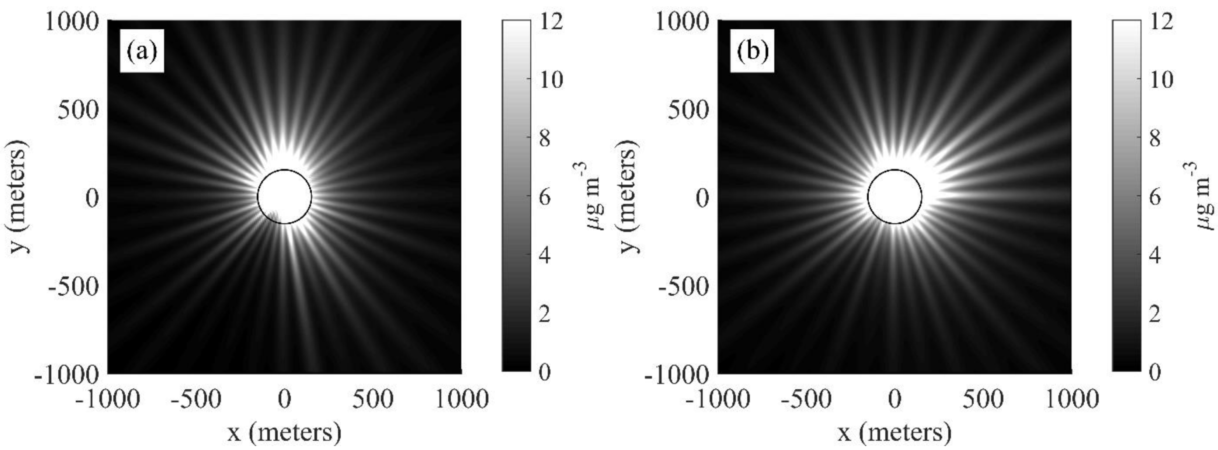
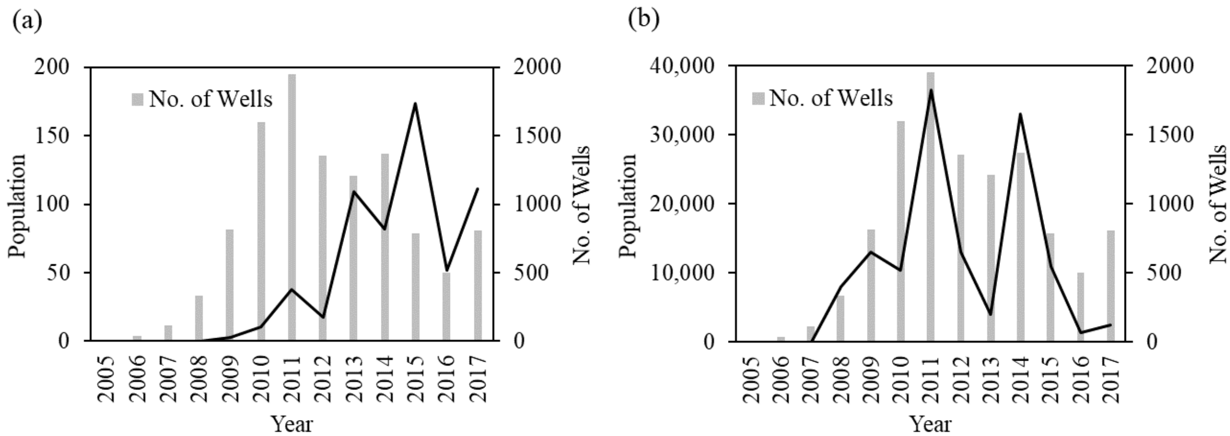


Figure-2: Map of PM concentrations originated from shale gas activities in Allegheny County, PA: a) a wellpad with 6 wells in 2017, and b) a wellpad with 12 wells in 2015. The black circle in both maps indicates the setback distance in Pennsylvania which is 152.4 m (500 ft.).

Figure-2a plots the changes in $PM_{2.5}$ concentration due to development activities on a wellpad with 6 wells in Allegheny County, PA, in 2017. Figure-2b presents the same concept for the case of a wellpad with 12 wells which was drilled in 2015 in the same county. Background concentration is not reflected in the depicted concentration values. Our model estimated that the number of people who could be exposed to concentrations higher than the standard in the latter case is zero, while it is at least 56 persons for the case of the wellsite with only 6 wells (panel (a)). The location of these two wellpads with respect to population is demonstrated in Figure-12S (refer to supplemental file).

Results from evaluation of affected population by $PM_{2.5}$ emissions from shale gas development at the county-level demonstrate the significance of proximity to residential areas as a deterministic factor. For instance, results show that in 2011, the number of additional exceedances from shale gas development to those attributable to background was estimated to be 1,183 persons in Butler County. The corresponding number of exceedances in Westmoreland County in 2012 as a result of emissions from development of similar number of wells to the previous case (43 versus 35 wells, respectively) was 10,128 persons. The number of exposure exceedances in the case of emissions from shale gas development only are zero in both counties in these two years which is expected considering the median number of wells per wellpad to be 1 in both cases (the mean was 1.75 and 1.95 in Butler and Westmoreland, respectively). Therefore, the probability of $PM_{2.5}$ concentration exceeding the annual limit beyond the setback distance from these wells would not be significant. However, additional exceedances would be expected when the level of background $PM_{2.5}$ concentrations and population densities within the vicinity of developed wells are take into account; corresponding average population densities were 492 and 881 persons/km^2 in Butler County (2011) and Westmoreland County (2012), respectively.



Year		2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Affected Population	Panel (a)	0.0	0.0	0.0	0.0	2.6	10.3	37.5	17.4	109.0	81.6	173.8	51.6	111.2
	Panel (b)	0.0	5.3	6.9	8059.8	12949.4	10400.6	36448.9	12955.8	4027.9	32970.6	10949.5	1379.7	2449.0
No. of Wells		9	36	112	331	815	1597	1953	1353	1208	1366	784	499	808
Average Background Conc. ($\mu g/m^3$)		15.3	13.3	14.5	12.6	11.3	11.9	11.2	10.4	10.2	10.3	10.0	8.7	9.5

Figure-3: Trend of annual number of populations experiencing concentrations higher than EPA's annual $PM_{2.5}$ standard due to shale development: (a) shale emissions only (zero background concentration), and (b) the additional exceedances occurring due to shale gas emissions from background only. Bars indicate the number of developed wells in each year.

Results from two main model output metrics of potentially affected people due to concentrations higher than standard are depicted in Figure-3. Figure-3a plots the time series of the number of people who might have experienced concentration levels higher than annual standard caused by emissions from shale gas development with a pristine background (i.e. no other $PM_{2.5}$). Figure-3b depicts the time series for the number of people who lived in areas where background concentration was lower than the standard, but the addition of shale gas development resulted in exceedance of the standard. More detailed data about the affected people by county and year is available in the supplemental file (Table-4S to Table-6S). Figure-3 demonstrate the results associated with the drilling rate of 1000 feet per day. The number of people who could experience exceedances due to shale gas development would be higher at slower daily rate of drilling. For example, in the case of shale emissions only (Figure-3a), 174 persons would increase to 571 persons in 2015. Results from a sensitivity analysis on drilling rate of 600 feet per day rather than 1000 feet per day is provided in Figure-7S.

According to Figure-3b, $PM_{2.5}$ emissions from shale gas development were the reason for exposure of more than 36,000 people to concentrations higher than annual standard in a single year (2011). That is approximately 0.8% of the Marcellus Shale regional population in Pennsylvania and equal to 3.25% of number of exposure exceedances occurring at ambient $PM_{2.5}$ (i.e., background only). This value does not include every possible person affected by Shale gas development in the corresponding year, as it leaves out those areas already non-compliant with the $PM_{2.5}$ standard.

The other notable trend in Figure-3 is the higher number of affected people in 2014 than in 2012, despite the similar number of developed wells. Table-1 provides the main statistics of wellpad density of all developed wells during 2005 to 2017 which demonstrate higher wellpad density in 2014 than in 2012. In accordance with data from Table-1, well data indicates that in 2012, 88% of the wellpads had 1 to 4 wells each, and the rest of them (12%) had 5 to 12 wells per wellpad. These numbers were 74% and 25%, respectively, in 2014. Also, the number of involved counties in these two years (31 in 2012 vs. 22 in 2014) indicates the possibility of more dispersed wellsites in 2012.

Combination of emissions from multiple sources can lead to $PM_{2.5}$ exposures higher than standard. Comparison of results from single and combined emissions of developed wellpads show that 48% of people experiencing PM concentrations higher than standard were due to emissions from multiple sites (Figure-11S).

Table-1: Wellpad density statistics. The 1st quartile, median, and 3rd quartile values represent 25th, 50th, and 75th percentiles, respectively.

Year	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Minimum	0	1	1	1	1	1	1	1	1	1	1	1	1
1st Quartile	0	1	1	1	1	1	1	1	1	1	1	2	2
Median	0	1	1	1	1	1	1	1	2	3	3	3	4
Mean	0	1.3	1.2	1.3	1.6	2.0	2.3	2.2	2.7	3.4	3.6	3.9	4.0
3rd Quartile	0	1	1	1	2	2	3	3	4	5	5	5	6
Maximum	0	5	4	13	23	27	22	12	11	19	21	15	12

Figure-4 demonstrates the decrease in the number of people who would have experienced concentrations higher than EPA standard due to shale gas development if the required setback distance had been greater (all else being equal). This figure presents the results for two main cases: 1) total repetition, and 2) no repetition. The case "total repetition" allows for maximum number of recurrent exceedances in different years. In other words, the person who experienced exposure exceedance in a specific year, also experienced this exceedance in all other years with equal or greater number of exceedance cases and in that regard, the case "total repetition" provides an estimation of a lower bound on the estimated number of people who experienced exposure exceedances from EPA annual limit due to shale gas emissions. In contrary, the case "no repetition" provides an estimation of reduction in exposure exceedances assuming that no person experienced exposure exceedance in more than one year and each case of exposure exceedance is a unique one. Therefore, the case "no repetition" estimates an upper bound on exposure exceedances. We use the unit of person-years for the total number of people who experienced concentration above the limit over the whole period of 2005 to 2017 in order to consider the possibility that same person could be counted in multiple years. Results indicate that increasing the setback distance from shale gas wellsites to 304.8 m (doubled), would reduce the number of exceedances by emissions only from shale developments by 77% (comparing to 174 person-years at current setback, in case of total repetition) to 95% (comparing to 595 person-years at current setback, in case of no repetition) (Figure-4a). More detail is available in the supplemental file (Figure-8S to Figure-10S).

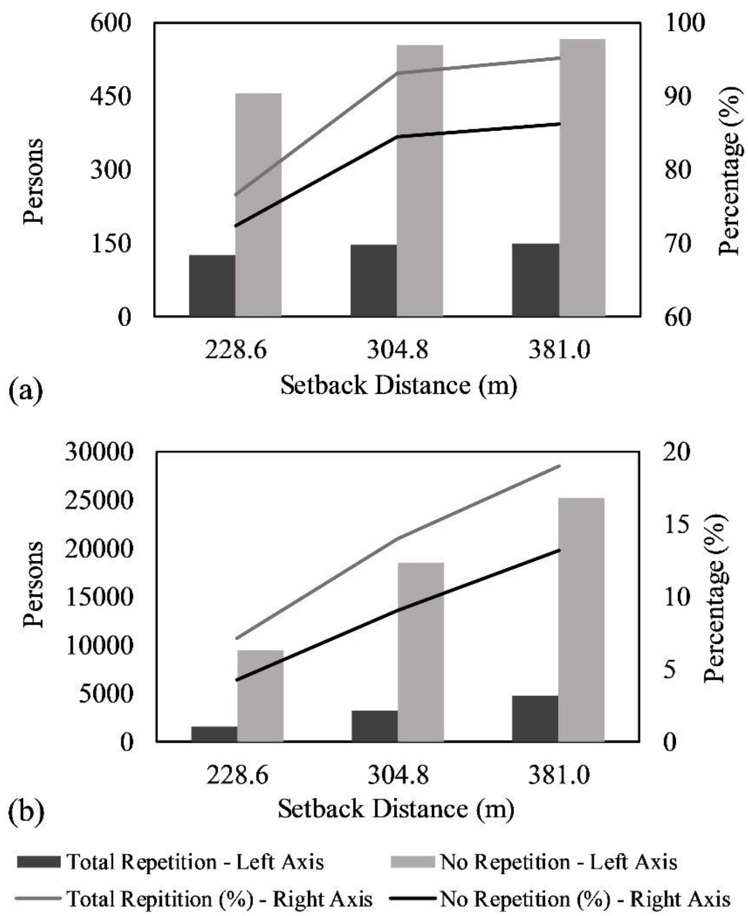


Figure-4: Reduction in the population exposed to $PM_{2.5}$ concentrations higher than the standard due to increase in setback distance from 152.4 m, in case of: (a) shale emissions only, and (b) the additional exceedances occurring due to shale gas emissions from background only. The black bars and lines are associated with the reduction in exposure exceedances while allowing for maximum number of recurrent exceedances in different years. The gray bars and lines are associated with the reduction in exposure exceedances assuming that each case of exposure exceedance is a unique person; i.e., no person experienced exposure exceedance in more than one year.

Figure-5 depicts the fractional reduction in the number of people who could have experienced $PM_{2.5}$ concentration higher than standard due to an increase to 381 m in setback distance. Results indicate that the majority of potential benefit from such an increase in the setback distance comes from a relatively small number of wellsites. Figure-5b demonstrates that 25% of the reduction (6263 person-years) in the number of additional exceedances from shale gas development to those attributable to background comes from only 10 wellsites (out of a total of 5109). Increasing the setback distance from 152.4 m. to 381 m. at only 54 wellsites (just 1.1% of the total wellsites or 2% of the total wells) would reduce the number of people experiencing $PM_{2.5}$ exceedances by 50% (12430 person-years).

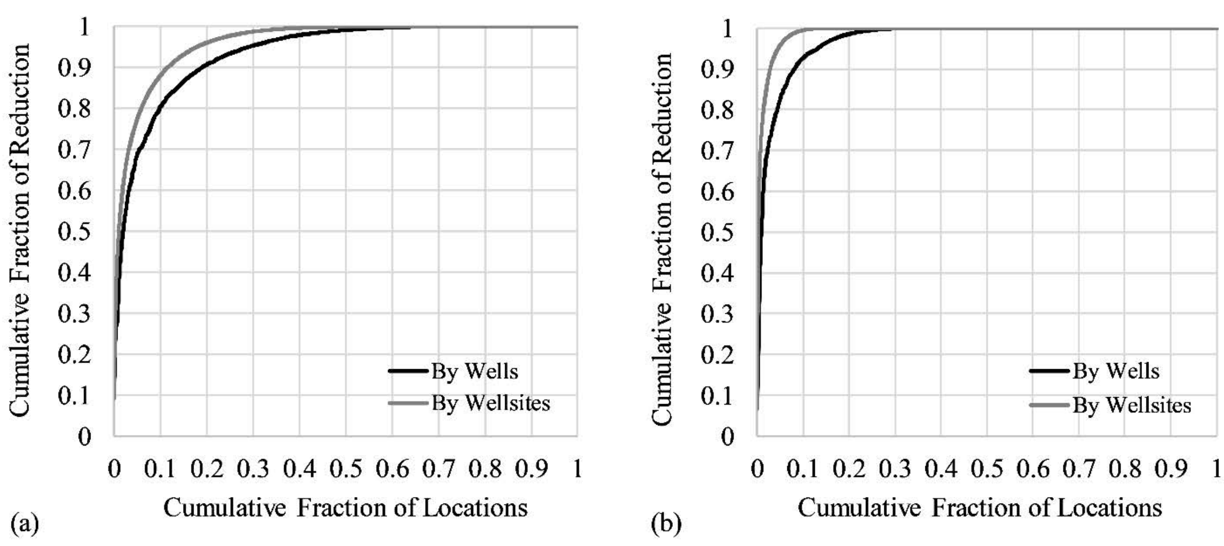


Figure-5: Cumulative fraction of reduction in $PM_{2.5}$ exposure exceedance by setback increase of 500 to 1250 feet in the case of: (a) shale emissions only, and (b) additional exceedances from shale gas development to those attributable to background.

According to Figure-5a, increasing the setback distance for the least impactful 68% of the wells and 81% of the wellsites would only achieve a collective industry-wide reduction of 5% in person-years exceedances of the $PM_{2.5}$ standard. However, increasing the setback distance would have no effect at all for 22% of the wells and 32% of the wellsites in reducing exceedances of the $PM_{2.5}$ standard (shale emissions plus background) (Figure-5b).

4. Discussion

Results indicate that policy surrounding shale gas development in Pennsylvania has limitations, as emissions from these activities in addition to existing background may have resulted in $PM_{2.5}$ concentrations exceeding the EPA's annual standard for up to 36,449 persons per single year between 2005 and 2017. This number represents 0.8% of the Marcellus Shale regional population in Pennsylvania (i.e., residents of the counties which were involved in development activities between 2005 and 2017) which represents a small portion. However, results indicated that corrective strategies even at a relatively small number of wellsites could potentially create major benefit (Figue-5). Moreover, our model estimated the average increase in level of $PM_{2.5}$ concentrations that led to exceedances to be 1.27 $\mu g/m^3$. This value is not very high in absolute terms and it emphasizes the important influence of shale gas development even when it causes small changes in overall air quality, especially those areas that are just barely meeting the current standard.

4.1. Setback policy should account for population and well development density

This analysis shows that there are real benefits to increasing the setback distance from developed wells, but that these benefits come from a relatively small set of wellsites and so such a change would not need to apply to all wells to have the desired effect. Fry (2013) argued that there was an absence of any technical basis in establishing the current setback policies, and this analysis supports this conclusion. To further limit negative air quality impacts of this industry, a more sophisticated set of policies is necessary to guarantee public health protections while allowing for continuing development of shale gas plays. Consistent with the discussion by Brown et al. (2015) regarding the dependence of associated health risk to emissions from shale gas development on time and location, these evaluations suggest that any setback and other exposure-limiting policy needs to consider wellpad density, well area-density and well locations as deterministic factors. Overall, wellpad density could be limited for wellpads located closer to populated areas and then relaxed as wellpads move away from these areas. Also, well area-density needs to be accounted for, specifically closer to more populated areas and those areas with background concentrations already approaching the standard. Such a rule could place an annual limit on the total number of wells within certain sub-divisions of a county. Such provisions could be accomplished more easily at the local level as the local governments could weigh and tradeoff the temporal and spatial density of these activities based on the population density, weather pattern, background concentration and the history of developments within the local vicinity of wells. Again, most wells due to their rural locations have minimal effect on residents' air quality in Pennsylvania, however, this is not true when development becomes very dense or is located close to populated areas. These situations should receive greater attention.

This analysis has examined changes to setback policy as if these changes could simply exclude population from a new radius around the existing wellsite during its development. In reality, this is unlikely to be the case. New setback policy applied to certain well development plans would likely result in a variety of effects including different land use patterns (e.g., more or fewer wells per pad), and longer drilling distances (i.e. each site would be emitting $PM_{2.5}$ for a longer period). It is difficult to anticipate all of these changes and their effects on local air quality, so future policy makers should prepare to be adaptable in mitigating these impacts.

4.2. Increasing wellpad density or well area-density increases affected population

Wellpad density seems to play a significant role in increasing emission concentrations within the vicinity of a wellsite (Figure-2) which was also discussed by Banan and Gernand (2018). As shown in Figure-2, higher wellpad density leads to more geographic area experiencing exceedance of the standard. Comparison of development in years 2012 and 2014 (Figure-3) indicates that such a case could cause an increase in the number of people who could be exposed to a hazardous level of concentrations from development of similar number of wells by a factor of 5. These results suggest that management of the air quality impacts of new shale gas wells needs to be managed collectively and not on a single site by site basis.

4.3. Well locations have moved closer to residential areas

Wellpad density and well area-density both need to be considered in discussing some of the cases where fewer people could have been affected by $PM_{2.5}$ emissions despite more wells being developed. However, these two do not exclusively govern $PM_{2.5}$ exposure outcomes. The role of well location is as significant as wellpad density and well area-density on the quantity of affected people by emission from shale gas development. Well location defines not only the differences in the distribution of wind speed and direction at different locations, but also the spatial distribution of nearby residential areas, and the existing background levels of $PM_{2.5}$.

During the period of shale gas development in Pennsylvania, the number of wells developed annually increased steadily through 2011 before decreasing in response to declining domestic natural gas prices. Well location maps indicate no obvious pattern in the changes to the distribution of wells (e.g., Figure-1) or locations of them in relation to major populated areas. An analysis of nearby populations versus well locations (Figure-4S and Figure-3) shows that existing policies and management actions, if any, were not successful in reducing exposure to $PM_{2.5}$ emissions. Even assuming a perfectly pristine background, the number of people who potentially experienced concentrations higher than the EPA's annual standard was as high as 174 people in 2015, assuming a high rate of emissions and the drilling rate of 1000 feet per day. According to the results, exposure exceedances with respect to corresponding emission rates are estimated to take the mean value of 15 and the 95% CI [0.31, 173.85] in 2013. See Figure-7S for the sensitivity analysis on a slower rate of drilling (600 feet per hour) and high hourly emission rates over the period of study.

Comparison of shale gas well locations and population data indicate that overall well development moved closer to more populated areas over the period of 2005 to 2017. We calculated the distance of any single well from the centroid of all the census blocks in Pennsylvania and summed the population of all the census blocks located within 1 km or less from that well. Our evaluation shows that for 5% of these wells, the population could be between 300 and 2100 people. A relatively small number of the wells, therefore, account for much of the exposure (Figure-6S).

It is important to also consider for population density within the vicinity of a well. The highest value of potentially affected people by shale gas development does not occur in the year wherein developed wells were located closest to residential areas (compare Figure-4S and Figure-3). Evaluation of well location versus population density helps to explain such a discrepancy. Assuming perfect enforcement of setback policy, results indicate that closest residential areas to development wellsites had higher population densities in 2017 than in 2011. In fact, the median values of population density in 2011 and 2017 were 12.9 and 15.6 persons per square-kilometer, respectively (refer to Table-1S). This highlights the important role of location in any investigation of shale gas health effects. Further analysis is needed to determine why development is moving closer to residential areas, whether it is due to improved information on available gas resources in these locations, or less resistance to development.

The actual placement of wells in practice is a compromise between the properties of the gas reservoir below the surface, the surface land available for lease or purchase, and the residential and environmental considerations of that land. While that decision is likely to remain a complex one, these results show that it is possible to locate seemingly large numbers of wells in a configuration that minimally affects residents' air quality, without reducing the overall activity levels of the industry.

This study simulated the annual concentration of $PM_{2.5}$ only, and not the other pollutants emitted from these sites. Therefore, these results do not provide any evaluation on sufficiency of current setback policy to meet EPA's standards on other pollutants, such as ozone and nitrogen oxides. The assumptions of the same hourly emission rate for all the engines used at development wellsites could lead to overestimation or underestimation of actual changes in $PM_{2.5}$ concentration due to each well. Also, the treatment of meteorological conditions and background concentrations result in some smoothing of actual exposures, along with the smoothing effects of the Gaussian plume model itself, which makes these estimates more reliable as indicators of long-term averages rather than specific historical exposures for any given area over short timescales. Rain and wet ground would reduce the PM concentrations. So, the assumption of dry ground in this evaluation is a conservative one made with the intent to lean towards greater protection of residents more than an accurate depiction of past exposure. A future analysis could incorporate precipitation to further enhance the accuracy of these results that do account for wind speeds and direction, cloud coverage, drilling depth, and estimated fracturing stages.

Accounting for the changes in the $PM_{2.5}$ concentrations at the location of schools, hospitals, daycares, or other important facilities would represent important estimations of vulnerability to exposure. However, such estimations require an access to data on operation timesheets as well as the activities' timing and timely records of residing population at these facilities which is not currently accessible in the available datasets.

Moreover, sometimes operators take advantage of fuel mixing by switching to natural gas for a portion of the engines' fuel as soon as the fracturing of the first well is complete and natural gas is being produced. Under such conditions, the emissions rate from fracturing of subsequent wells of the same wellpad would be lower. However, data on the frequency and extent of this practice are not available. So, the estimated concentrations are likely overestimated to some extent for the wellsites engaging in this practice, but the degree to which this is true cannot be known at this time. Despite these limitations, this study is the first attempt to estimate the total number of people experiencing exceedance of the $PM_{2.5}$ standard due to emissions from shale gas development activities.

5. Conclusion

The current study evaluates the efficacy of current setback policy according to actual shale gas developments over the period of 2005 to 2017 by adding temporal and spatial details to this picture. Results from this study support previous arguments on the inadequacy of setback policy in Pennsylvania (152.4 m from a wellpad) for avoiding exceedances of the EPA's annual limit on $PM_{2.5}$. These findings indicate a proportional increase in the number of affected residents in association with a greater number of developed wells. Meanwhile, previous development histories in some counties and in some years show that similar levels of development could occur with reduced population exposure.

Overall, results indicate that increasing numbers of wells have been accompanied by increases of at least one of these three variables: wellpad density, well area-density, and proximity to residential areas. However, none of these deterministic factors, i.e. wellpad density, well area-density and activity location, govern this exceedance alone. It is the combined effect of these factors which explains the trend of changes in the number of affected people versus number of developed wells. Rather than blanket changes to the setback policy to be applied statewide, there is substantial potential to achieve real benefits in terms of reduced exposure by focusing on a small number of development sites and taking steps to push those sites further from populated areas.

This analysis sheds light on the significant and entangled effect of wellpad density, well area-density, and well locations on local air quality, with respect to concentration of $PM_{2.5}$, and justifies the need for a technical basis in evaluating the setback policy and any other related limitations on the development of shale gas wells. These results indicate that well location is equally important as wellpad density or well area-density, and the effect of each varies based on population density and weather pattern within the vicinity of each shale gas development site. Under the most rigorous consideration, setback policy could be designed based on the common distance from more populated areas, aiming to set a cap on the local wellpad density or well area-density. This policy could conditionally be relaxed for the development cases occurring within areas with limited or no population.

Results from this study plus the stricter standards on other associated pollutants, such as NOx and VOCs, emphasize the necessity for the effects of other pollutants to be investigated. Sensitivity analyses in this study indicated a crucial need for expanded and continued direct monitoring of air quality in this region at a higher spatial resolution that would support improved environmental management of this diffuse industry. The same argument is valid for more precise meteorological data at ground level at the location of each wellpad. Results from this study have demonstrated the importance of accuracy in data pertaining to emission rates from engines, local meteorological data, background concentration, timing of shale gas development activities, and operational details as fundamental inputs in simulation of hazardous emissions dispersion from shale gas developments. The sensitivity of the outputs to accuracy of these data emphasizes the need for entities, such as Department of Environmental Protection, to set regulatory obligations on reporting detailed operational data by operators as well as allocation of financial resources to development or improvement of datasets (including emissions rates, emission measurements and background concentrations) to reduce the uncertainties in this kind of impact assessment.

Acknowledgements

The authors are grateful to Professor Mort D. Webster and Dr. Eugene Morgan for providing intellectual and material resources to improve this study.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

6. References

1. Adgate, J.L., Goldstein, B.D., McKenzie, L.M., 2014. Potential Public Health Hazards, Exposures and Health Effects from Unconventional Natural Gas Development. Environmental Science and Technology. 48, 8307-8320.

2. Amoatey, P., Omidvarborna, H., Affum, H.A., Baawain, M., 2019. Performance of AERMOD and CALPUFF models on SO_2 and NO_2 emissions for future health risk assessment in Tema Metropolis. Human and Ecological Risk Assessment: An International Journal. 25(3): 772-786.

3. Arystanbekova, N.K., 2004. Application of Gaussian plume models for air pollution simulation at instantaneous emissions, Mathematics and Computers in Simulation. 67, 451-458.

4. Banan, Z., Gernand, J.M., 2018. Evaluation of gas well setback policy in the Marcellus Shale region of Pennsylvania in relation to emissions of fine particulate matter. Journal of the Air & Waste Management Association. 68:9, 988-1000.

5. Brown, D., Weinberger, B., Lewis, C., Bonaparte, H., 2014. Understanding exposure from natural gas drilling puts current air standards to the test. Reviews on Environmental Health. 29(4): 277-292.

6. Brown, D.R., Lewis, C., Weinberger, B.I., 2015. Human exposure to unconventional natural gas development: A public health demonstration of periodic high exposure to chemical mixtures in ambient air. Journal of Environmental Science and Health. Part A, 50:5, 460-472.

7. Census Data, 2010. https://www.census.gov/geo/maps-data/data/tiger-data.html; (last accessed on January 23, 2019)

8. Cheng, H., Small, M.J., Pekney, N.J., 2015. Application of nonparametric regression and statistical testing to identify the impact of oil and natural gas development on local air quality. Atmospheric Environment. 119, 381-392.

9. Coloradans for Responsible Energy Development, 2019. https://www.cred.org/fracking-frequency/; (last accessed on May 25, 2019).

10. Croft, D.P., Zhang, W., Lin, S., Thurston, S.W., Hopke, P.K., Masiol, M., Squizzato, S., Wijngaarden, E., Utell1, M.J., Rich, D.Q., 2019. The association between respiratory infection and air pollution in the setting of air quality policy and economic change. Annals of the American Thoracic Society. 16(3): 321-330.

11. Department of Oil and Gas Reporting website, 2018. http://www.depreportingservices.state.pa.us/ReportServer/Pages/ReportViewer.aspx?/Oil_Gas/Permits_Issued_Detail; (last accessed on November 8, 2018)

12. Drillinginfo, 2019. https://info.drillinginfo.com/; (last accessed on May 20, 2019).

13. EIA – Shale gas production, 2018. http://www.eia.gov/dnav/ng/ng_prod_shalegas_s1_a.htm; (last accessed on November 8, 2018)

14. EPA, 1995. User's Guide for the Industrial Source Complex (ISC3) Dispersion Models: Volume II – Description of Model Algorithms. U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards – Emissions, Monitoring, and Analysis Division, Research Triangle Park, North Carolina.

15. EPA, 2018. https://www.epa.gov/outdoor-air-quality-data/download-daily-data; (last accessed on November 5, 2018).

16. EPA Oil and Gas Dictionary, 2018. http://files.dep.state.pa.us/OilGas/BOGM/BOGMPortalFiles/OilGasReports/HelpDocs/SSRS_Report_Data_Dictionary/DEP_Oil_and_GAS_Reports_Data_Dictionary.pdf; (accessed May 31, 2019).

17. Facts about Canada's Oil and Natural Gas Industry, 2019. https://oilandgasinfo.ca/know-fracking/how-long-does-fracking-take/ ; (last accessed on May 20, 2019).

18. Fry, M., 2013. Urban gas drilling and distance ordinances in the Texas Barnett Shale. Energy Policy. 62, 79-89.

19. Gilman, J.B., Lerner, B.M., Kuster, W.C., de Gouw, J.A., 2013. Source signature of volatile organic compounds from oil and natural gas operations in northeastern Colorado. Environmental Science and Technology. 47(3): 1297-1305.

20. Gopoch, A., Burk, S., Davidson, K.L., 1980. Stability effects on aerosol size and height distributions. Tellus. 32:3, 245-250.

21. Haley, M., McCawley, M., Epstein, A.C., Arrington, B., Bjerke, E.F., 2016. Adequacy of Current State Setbacks for Directional High-Volume Hydraulic Fracturing in the Marcellus, Barnett, and Niobrara Shale Plays. Environmental Health Perspectives. 124:9, 1323-1333.

22. ICF International, 2009. Well Permit Issuance for Horizontal Drilling and High-Volume Hydraulic Fracturing to Develop the Marcellus Shale and Other Low Permeability Gas Reservoirs. ICF International Incorporated, LLC for NYSERDA: Albany, NY.

23. IEM: Iowa Environmental Mesonet, 2018. http://mesonet.agron.iastate.edu/request/download.phtml?network=PA_ASOS; (last accessed on November 8, 2018)

24. Jemielita, T., Gerton G.L., Neidell M., Chillrud, S., Yan, B., Stute, M., Howarth, M., Saberi, P., Fausti, N., Penning, T.M., Roy, J., Propert, K.J., Panettieri, R.A., 2015. Unconventional

Gas and Oil Drilling is Associated with Increased Hospital Utilization Rates. PLOS ONE. 10:7, e0131093.

25. Kelly F.J., Fussell, J.C., 2012. Size, source and chemical composition as determinants of toxicity attributable to ambient particulate matter. Atmospheric Environment. 60, 504-526.

26. Krzyzanowski, J., 2012. Environmental pathways of potential impacts to human health from oil and gas development in northeast British Columbia, Canada. Environmental Reviews. 20(2): 122-134.

27. Martin, A.R., Finlay, W.H., 2005. The Effect of Humidity on the Size of Particles Delivered from Metered-Dose Inhalers. Aerosol Science and Technology. 39:4, 283-289.

28. McKenzie, L.M., Allshouse, W.B., Burke, T., Blair, B.D., Adgate, J.L., 2016. Population Size, Growth, and Environmental Justice near Oil and Gas Wells in Colorado. Environmental Science and Technology. 50, 11471-11480.

29. Matacchiera, F., Manes, C., Beaven, R.P., Rees-White, T.C., Boano, F., Mønster, J., Scheutz, C., 2019. AERMOD as a Gaussian dispersion model for planning tracer gas dispersion tests for landfill methane emission quantification. Waste Management. 87, 924-936.

30. McKenzie, L.M., Witter, R.Z., Newman, L.S., Adgate, J.L., 2012. Human health risk assessment of air emissions from development of unconventional natural gas resources. Science of the Total Environment. 424, 79-87.

31. McKeon, M., 2011. Horizontal Fracturing in Shale Plays (Workshop Presentation). Petroleum Transfer Technology Council. http://www.thepttc.org/workshops/eastern_062111/eastern_062111_McKeon.pdf; (last accessed on May 20, 2019).

32. Meng, Q., 2015. Spatial analysis of environment and population at risk of natural gas fracking in the state of Pennsylvania, USA. Science of the Total Environment 515-516, 198-206.

33. Meng, Q., Ashby, S., 2014. Distance: A critical aspect for environmental impact assessment of hydraulic fracking. The Extractive Industries and Society. 1, 124-126.

34. Moore, C.W., Zielinska, B., Petron, G., Jackson, R.B., 2014. Air impacts of increased natural gas acquisition, processing, and use: a critical review. Environmental Science and Technology. 48, 8349-8359.

35. Petersen, R.L., Guerra, S.A., Bova, A.S., 2017. Critical review of the building downwash algorithms in AERMOD. Journal of the Air & Waste Management Association. 67:8, 826-835.

36. Rabinowitz, P.M., Slizovskiy, I.B., Lamers, V., Trufan, S.J., Holford, T.R., Dziura, J.D., Peduzzi, P.N., Kane, M.J., Reif, J.S., Weiss, T.R., Stowe, M.H., 2015. Proximity to Natural Gas Wells and Reported Health Status: Results of a Household Survey in Washington County, Pennsylvania. Environmental Health Perspectives. 123:1, 21-26.

37. Rasmussen, S.G., Ogburn, E.L., McCormack, M., Casey, J.A., Bandeen-Roche, K., Mercer, D.G., Schwartz, B.S., 2016. Association between Unconventional Natural Gas Development in the Marcellus Shale and Asthma Exacerbations. JAMA Internal Medicine. 176(9):1334-1343.

38. Roy, A.A., Adams, P.J., Robinson, A.L., 2014. Air pollutant emissions from the development, production, and processing of Marcellus Shale natural gas. Journal of the Air & Waste Management Association. 64:1, 19-37.

39. Sinclair, D., Countess, R.J., Hoopes, G.S., 1974. Effect of Relative Humidity on the Size of Atmospheric Aerosol Particles. Atmospheric Environment. 8, 1111-1117.

40. Tartakovsky, D., Broday, D.M., Stern, E., 2013. Evaluation of AERMOD and CALPUFF for predicting ambient concentrations of total suspended particulate matter (TSP) emissions from a quarry in complex terrain. Environmental Pollution. 179, 138-145.

41. Touma, J.S., Isakov, V., Ching, J., Seigneur, C., 2006. Air Quality Modeling of Hazardous Pollutants: Current Status and Future Directions. Journal of the Air & Waste Management Association. 56:5, 547-558.

42. Turner, D. B., 1970. Workbook of Atmospheric Dispersion Estimates. Public Health Service Publication No. 999-AP-26, U.S. Department of Health, Education and Welfare, National Air Pollution Control Administration, Cincinnati, Ohio.

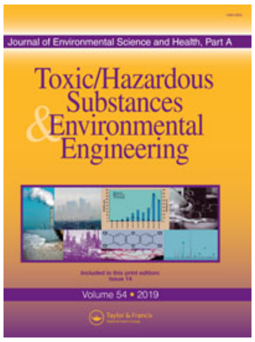
43. WHO, 2007. Health relevance of particulate matter from various sources. Report on a WHO workshop, March 2007. World Health Organization: Bonn, Germany.

44. Winkler, P., 1988. The Growth of Atmospheric Aerosol Particles with Relative Humidity. Physica Scripta. 37, 223-230.

45. Zielinska, B., Fujita, E., Campbell, D., 2011. Monitoring of Emissions from Barnett Shale Natural Gas Production Facilities for Population Exposure Assessment. Mickey Leland National Urban Air Toxics Research Center

ATTACHMENT C

STUDY 20



Journal of Environmental Science and Health, Part A
Toxic/Hazardous Substances and Environmental Engineering

ISSN: 1093-4529 (Print) 1532-4117 (Online) Journal homepage: www.tandfonline.com/journals/lesa20

Assessing exposure to unconventional natural gas development: using an air pollution dispersal screening model to predict new-onset respiratory symptoms

David R. Brown, Lydia H. Greiner, Beth I. Weinberger, Leslie Walleigh & Dale Glaser

JOURNAL OF ENVIRONMENTAL SCIENCE AND HEALTH, PART A
2019, VOL. 54, NO. 14, 1357–1363
https://doi.org/10.1080/10934529.2019.1657763

Assessing exposure to unconventional natural gas development: using an air pollution dispersal screening model to predict new-onset respiratory symptoms

David R. Brown[a], Lydia H. Greiner[a,b], Beth I. Weinberger[a], Leslie Walleigh[a], and Dale Glaser[b]

[a]Southwest Pennsylvania Environmental Health Project, McMurray, PA, USA; [b]San Diego State University, San Diego, CA, USA

ABSTRACT

Various exposure estimates have been used to assess health impact of unconventional natural gas development (UNGD). The purpose of this study was to (1) use an air pollution dispersal screening model and wind direction to characterize the air emissions from UNGD facilities at each residence and (2) assess association of this exposure estimate with respiratory symptoms. Respiratory symptoms were abstracted from health records of a convenience sample of 104 adults from one county in southwestern PA who had completed a standard clinical interview with a nurse practitioner. Using publicly available air emission data, we applied a "box" air pollution dispersion screening model to estimate the median ambient air level of CO, NO_x, PM 2.5, VOCs, and formaldehyde at the residence during the year health symptoms were reported. Sources and median emissions were categorized as north, south, east, or west of the residence to account for the effect of wind direction on dispersion. Binary logistic regression was performed for each respiratory symptom. Number of sources had varying magnitudes of association with some symptoms (i.e., cough, shortness of breath, and "any respiratory symptom") and no association with others (i.e., sore throat, sinus problems, wheezing). Air emissions were not associated with any symptom.

ARTICLE HISTORY
Received 6 May 2019
Accepted 7 August 2019

KEYWORDS
Hydraulic fracking; unconventional natural gas; air emissions; exposure; health

Introduction

Over the past decade many areas of rural Pennsylvania have become industrialized by the rapid expansion of unconventional natural gas development (UNGD). This development has resulted in thousands of point sources of hazardous air emissions distributed over the previously rural landscape. In spite of the well documented harmful health effects of some of the most abundant emissions from UNGD, such as particulate matter, nitrogen oxides, and volatile organic compounds,[1,2] the overall public health impact of UNGD has been difficult to assess due to limited exposure information.

Epidemiologic studies have looked at the associations between UNGD exposure and birth outcomes;[3–5] childhood cancer;[6] a constellation of symptoms commonly reported: headache, fatigue, and rhinosinusitis;[7] vehicular accidents;[8] depressive symptoms;[9] and asthma exacerbations.[10] Self-report studies have documented an association between respiratory symptoms and proximity to UNGD activity.[11]

Our previous work described the prevalence of self-reported symptoms in a sample of adults who lived within two kilometers of at least one unconventional natural gas well in southwestern PA.[12] Five of the 10 most frequently reported symptoms were respiratory: sore throat (39%); cough (33%); shortness of breath (29%); sinus problems (29%); and wheezing (22%). This self-reported prevalence of respiratory symptoms in residents of communities exposed to UNGD is consistent with the work of Rabinowitz et al.,[11] Rasmussen et al.,[10] and Tustin et al.[7] Rabinowitz et al. found a relationship between respiratory symptoms and increased proximity to UNGD wells; Rasmussen et al. and Tustin et al. found relationships between exposure and asthma exacerbations and sinusitis, respectively, using an exposure metric incorporating four stages of well development (pad preparation, drilling, stimulation, and production) along with the sum of the inverse squared distances between wells and residences.

One of the challenges of looking at health outcomes related to exposure is that it is unclear how best to characterize residential exposure to UNGD emissions. Different metrics, virtually all of which have focused on wells while ignoring potential contribution of other infrastructure, have been used to estimate exposure. Among the simplest is distance to unconventional gas wells.[11,13] Others have used the inverse distance weighted well count to account for wells over a large distance, giving greater weight to those closest to the residence.[3,4] Responding to the variation in emissions at different stages of well development, still others have incorporated those stages into metrics that included distance and number and depth of wells, and production of gas.[5,7,9,10] Koehler et al.[14] note wells are just a part of the infrastructure of UNGD. Additional sources of exposure include impoundments, where hydraulic fracturing fluid that returns to the surface from the well is stored; flaring events, where excess gas is burned off; compressor stations, that

CONTACT Lydia Greiner lgreiner@environmentalhealthproject.org Southwest Pennsylvania Environmental Health Project, 2001 Waterdam Plaza Drive, Suite 201, McMurray, PA 15317, USA.

keep the gas moving through the pipelines; and pipelines themselves, that carry the gas from processing plants to the end user. Underscoring the importance of including as many recognized sources of air emissions as feasible, they found that a metric that incorporated both proximity to four stages of well development and, additionally, to compressor stations was more predictive of mild asthma exacerbations than either of the two other metrics used in prior epidemiological studies (i.e., inverse weighted distance or proximity to four stages of well development). Although Koehler et al. initially intended to incorporate data on flaring events and impoundments, they found that the data on these two important sources of air emissions were too sparse.

We have developed two measures of exposure for this analysis to better capture sources of emissions and the impact of air movement on exposure. The goals of this analysis were to (1) characterize the UNGD-related emissions from well pads, processing plants, and compressor stations; (2) use an air pollution dispersal screening model and wind direction to estimate household-level exposure to these emissions; and (3) test the relationship between the exposure measures and respiratory health outcomes.

Materials and methods

Health outcomes

As previously described in detail,[12] the Southwest Pennsylvania Environmental Health Project has been systematically collecting health data from residents of communities located near UNGD sites since 2012. Between February 1, 2012 and December 31, 2017, 164 adults and children completed the standardized health assessment, typically conducted face-to-face by a family nurse practitioner using standard clinical practice for collecting current problems, review of systems, past medical history, family history, and social history. These individuals self-selected to complete the health assessment with the nurse practitioner because of concerns about symptoms they were experiencing. The 164 records in this convenience sample were reviewed retrospectively by a team of health care providers that included a physician who is board certified occupational medicine (LW) and at least one nurse practitioner.

For this analysis, exclusion criteria included: age less than 18; employment in the oil or gas industry; incomplete health assessment; and residence outside of the county of interest. After exclusion criteria were applied, a convenience sample of 104 records was available for this analysis. Symptoms were abstracted from each record; symptoms were excluded if they could plausibly be attributed to pre-existing or current health conditions (e.g., chronic obstructive pulmonary disease in the case of "shortness of breath") or behaviors (e.g., tobacco smoking in the case of "cough"). Five respiratory symptoms were abstracted and used as dichotomous outcome in this analysis: cough, shortness of breath, sinus problems, sore throat, and wheezing. A sixth dichotomous outcome, "any respiratory symptom", was used to indicate the presence of at least one respiratory symptom.

Exposure measures

Since March 31, 2012, owners and operators of natural gas production and processing operations have been required to report air emissions to the Pennsylvania Department of Environmental Protection (PA DEP) annually. Initially, owners and operators were required to report emissions from the following sources: stationary engines, heaters, tanks/impoundments, dehydration units, pneumatic pumps, fugitives, venting and blowdown, drill rigs, and well completions; in 2013, compressor stations were added to the list of sources. Air contaminants reported from these sources include: carbon monoxide, nitrogen oxides, particulate matter less than 10 micrometers in diameter (PM10), particulate matter less than 2.5 micrometers in diameter (PM2.5), sulfur dioxide, volatile organic compounds, and additional hazardous air pollutants including benzene, ethylbenzene, formaldehyde, *N*-hexane, toluene, and 2,2,4-trimethylpentane. Greenhouse gas emissions are also reported.[15] The publicly available annual Air Emission Inventory reports "tons per year" for each compound released from each source.[16] It is important to note that, while emissions from each type of source have some shared characteristics, they also differ in kind and quantity.[17]

Data from 2012–2016 were available.[16] We used the annual Air Emissions Inventory that corresponded to the year the health assessment was collected. Data from the Air Emissions Inventory that corresponded to the year of the health assessment were not available for 17 of the 104, reducing our sample to 87. Geocoding was used to determine the proximity of the sources listed in the Air Emissions Inventory to each residence. Latitude and longitude coordinates for each source are published in the Inventory. Addresses, available for each residence, were converted into latitude and longitude.[18] Using these coordinates, we identified all sources reported on the annual Air Emissions Inventory that were located within two kilometers of each residence. For our exposure estimate, we selected the five compounds with the highest reported mass and known health effects: nitrogen oxides, carbon monoxide, VOCs, PM 2.5, and formaldehyde. While the greenhouse gasses methane and carbon dioxide are among those compounds with the highest reported mass they were not included as they do not have acute health effects at levels typical for environmental exposures. Emissions data were reported for each compound as tons per year. For this analysis we converted tons/year into grams per hour.

There are limitations to the PA DEP data. Not all sources of emissions are included in the report. Emissions from the largest Title V compressor stations and selected compressor stations along interstate pipelines are reported only to the federal government and as a result are not included in the Air Emissions Inventory. Air emissions resulting from flaring events or evaporation from impoundments are not included in the Air Emissions Inventory, and indoor exposures to harmful compounds off gassed from contaminated water remain the subject of conjecture. Air emissions are reported as "tons per year"; our conversion to "grams per

Table 1. Air stability classes.[19]

Wind Speed (mph)	Day Clear or just a few clouds	Day <50% cloud cover	Day >50% cloud cover	Day >80% cloud cover	Night <50% cloud cover	Night >50% cloud cover
<5	A1	AB6	B11	D16	F26	E21
5–7	AB2	B7	C12	D17	F27	E22
7–11	B3	BC8	C13	D18	E28	D23
11–13	C4	CD9	D14	D19	D29	D24
>13	C5	D10	D15	D20	D30	D25

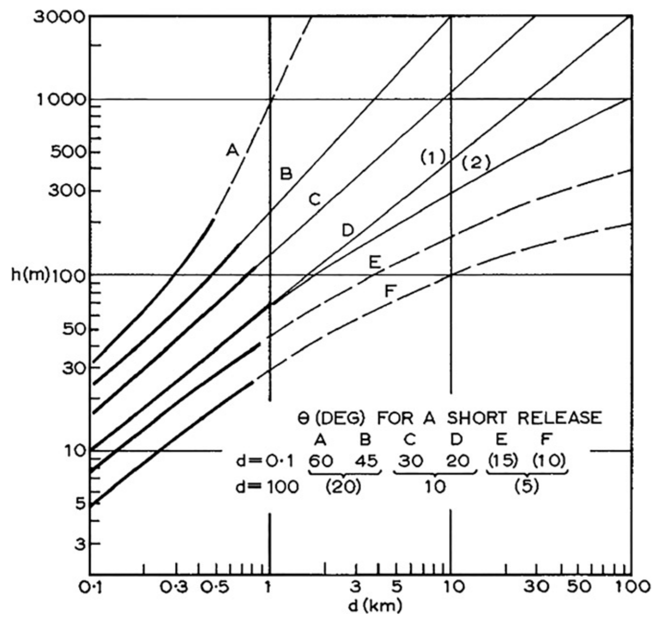


Figure 1. Vertical mixing by stability class and distance from source.[19]

hour" assumes a consistent rate of emissions and dilutes the potential effect of peak exposures.

Patterns of dispersal of air emissions are influenced by weather and atmospheric conditions. We estimated exposure at a residence using a "box" air pollution dispersion screening model, based on the work of Pasquill.[19,20] Our approach, described in detail elsewhere,[17,21] will be described briefly here. The "box" air pollution dispersion screening model is based on a theoretical box (volume of air) that carries emissions from a source. That box increases in size (dispersion) based on parameters that determine air dilution down-wind from emission sources to estimate the ambient air level of pollutants. The parameters are: (1) cloud cover which determines vertical mixing due to heating and cooling of ground surface; (2) wind speed which determines time for horizontal mixing; and (3) time of day, which influences stability of the air system. Using these parameters, Pasquill developed a model with six stability classes defined by particular combinations of atmospheric conditions, as shown in Table 1.

The initial volume of the box is calculated by

$$\text{Volume }(\text{m}^2) = a * b * c$$

where a = meters of air that pass over a site/minute; $b = 100$ meters (assumed dimension of typical site); and c = intercept of stability class and distance from source (see Fig. 1).

The air concentration of a compound at the source is calculated by

$$S(\mu\text{gm}^{-3}) = \text{Emissions/Volume}$$

where Emissions = mass in μg min^{-1} and Volume=a^*b^*c.

To determine the dilution downwind from the source the volume of the box at specific distances from the source must be recalculated by

$$\text{VOLUME }(\text{m}^2) = a * B * C$$

where a = meters of air that pass over a site/minute; B = expansion of the box (see Fig. 1); and C = intercept of stability class and distance from source (see Fig. 1).

We applied this model using data from the National Oceanic and Atmospheric Association (NOAA).[22] NOAA provides hourly cloud cover, wind speed, time of day, and wind direction. For this analysis, we used data collected at the Allegheny County Airport in West Mifflin, PA. This station is located on average 20–30 miles from the residences included in this analysis. Data from the West Mifflin station were used to establish hourly air stability classes over the year at each residence in our sample.

For each source within two km of a residence, we used the screening model to calculate the hourly ambient air levels at the source and at the residence in μg m^{-3} for CO, NO_x, PM 2.5, VOCs, and formaldehyde emitted over the year. We determined the mean, standard deviation, and median of the hourly air emissions for each compound and then summed the medians. Table 2 shows the results of this screening model applied to a hypothetical source emitting 300 g h^{-1}. As shown on Table 2, we used the model to calculate ambient air levels at seven unique distances ranging from 0.1 km (i.e., "fence line") to 10 km from the hypothetical source, with air emissions expressed in μg m^{-3}.[20]

The "box" air pollution dispersion screening model does not take in to account wind direction. Our examination of data from the National Oceanic and Atmospheric Association[22] revealed that wind direction is from the north, south, and west approximately 90% of the time (i.e., 30% of the time from each of these directions), and from the east approximately 10% of the time. To account for the variation in wind, each source was categorized as north, south, east, or west of the residence, based on distance from "true". For example, we considered a source north of a residence if it was within 45° of true north, and the same for the other three directions.

As an estimate of exposure to air emissions we generated eight variables for this analysis. For each residence we generated four variables ("north sources", "south sources", "east sources", and "west sources") which represented the number of sources in each quadrant within two km of the residence.

Table 2. Ambient air levels in $\mu g\ m^{-3}$ from a hypothetical source emitting 300 grams per hour by stability class.

	Distance from Source						
Stability Class	≤0.1 km	>0.1 km, ≤0.5 km	>0.5 km, ≤1 km	>1 km, ≤2 km	>2km, ≤3 km	>3km, ≤5 km	>5km, ≤10 km
A1	175	5	1	<1	<1	<1	0
AB2	125	8	2	<1	<1	<1	0
B3	75	8	2	<1	<1	<1	0
C4	125	22	7	2	1	<1	<1
C5	100	19	6	1	<1	<1	<1
AB6	175	12	3	<1	<1	<1	0
B7	150	13	4	<1	<1	<1	0
BC8	125	17	6	1	<1	<1	<1
CD9	175	28	10	3	1	<1	<1
D10	200	29	10	4	2	<1	<1
B11	200	18	6	1	<1	<1	<1
C12	250	41	14	4	1	<1	<1
C13	150	25	9	2	1	<1	<1
D14	225	33	12	4	2	<1	<1
D15	200	29	10	4	2	<1	<1
D16	625	87	32	12	6	2	<1
D17	450	62	23	9	4	1	<1
D18	275	39	14	5	2	1	<1
D19	225	33	12	4	2	<1	<1
D20	200	29	10	4	2	<1	<1
E21	875	150	73	35	21	10	3
E22	625	100	52	25	15	7	2
D23	275	39	14	5	3	1	<1
D24	225	33	12	4	2	<1	<1
D25	200	29	10	4	2	<1	<1
F26	1,400	225	100	56	33	15	5
F27	1,000	150	83	40	23	11	4
E28	375	78	33	16	9	4	1
D29	225	33	12	4	2	<1	<1
D30	200	29	10	4	2	<1	<1

Table 3. Tons per year of carbon dioxide, methane, nitrogen oxides, carbon monoxide, volatile organic compounds (VOCs), PM2.5, and formaldehyde emitted in Washington County PA 2012–2015.

Compound	2012	2013	2014	2015	2016
Carbon dioxide	681,402	731,097	897,398	1,028,463	1,293,388
Methane	9,204	7,217	3,069	7,684	8,276
Nitrogen oxides	2,097	2,875	3,931	4,546	3,544
Carbon monoxide	939	939	1,375	2,087	1,742
VOCs	639	1,658	1,548	1,932	2,078
PM2.5	66	102	157	111	85
Formaldehyde	54	55	55	61	63

We generated four additional variables ("north emissions", "south emissions", "east emissions", and "west emissions") which represented the median of ambient air levels of emissions from wells, processing plants, and compressor stations in the quadrant.

Analytic plan

Given the nature of the dichotomous outcomes (i.e., any respiratory symptom, cough, shortness of breath, sore throat, sinus problems, and wheezing), binary logistic regression was performed.[23] The quantitative predictors were four emissions (i.e., north, south, east, and west EM) and four sources (i.e., north, south, east, and west source). Moreover, gender, age, and household water source were included in the full model. All parameter estimates (e.g., the logit b, odds ratio (OR) and 95% confidence interval, etc.) are reported ($\alpha = .05$), as well as classification statistics such as sensitivity (true positive hit rate) and specificity (true negative hit rate). As well, pseudo r^2 statistics are reported, including the Cox–Snell and Nagelkerke R^2, each of which is a function of the –2LL statistic for the full and restricted model (LL = log likelihood). The Hosmer–Lemeshow test was used to assess model fit[24] and for this test statistic, non-significance is preferred (i.e., expected probabilities approximate observed probabilities).

Results and discussion

The median age of the 87 adults in this convenience sample was 57 (SD 12); 60% were female; and 40% reported using well water for human activities such as cooking, drinking, and/or bathing.

UNGD-related emissions from well pads, processing plants, and compressor stations

There are 16 compounds included on the Annual Emissions Inventory for gas wells and related facilities.[16] On Table 3, we show the tons/year of seven compounds emitted from well pads, processing plants, and compressor stations in Washington County, PA. These seven compounds are shown because they are consistently emitted in larger mass than other compounds.

Household-level exposure to emissions

Household level exposure in this sample varied by year and location of the source relative to the residence. Table 4

Table 4. Median ambient air levels of emissions of nitrogen oxides, carbon monoxide, VOCs, PM 2.5, and formaldehyde by year.

Year	North	East	South	West
2012	0	0	0	25
	(0,1562.16)	(0,999.36)	(0,3972.58)	(0,1556.67)
2013	93.51	145.2	206.6	268.0
	(0,945.5)	(0,1208.8)	(0,6638.3)	(0,761.8)
2014	0	7.96	0	5.80
	(0,1301.24)	(0,3619.44)	(0,3902.76)	(0,8670.90)
2015	197.03	0.09	238.44	307.16
	(0,3784.57)	(0,4542.18)	(0,2684.26)	(0,2169.16)
2016	312.9	73.00	327.4	460.0
	(0,6065.00)	(0,1027.0)	(0,3661.6)	(0,3057.0)

Table 5. Associations between exposure measures and any reported respiratory symptom.

	B	S.E.	*P*-value	OR	CI Lower	CI Upper
North EM	0	0	0.795	1	0.999	1.001
East EM	0	0	0.263	1	0.999	1
South EM	0.001	0.001	0.147	1.001	1	1.002
West EM	0	0	0.202	1	0.999	1
North Sources	−0.259	0.147	0.078	0.772	0.578	1.03
South Sources	−0.464	0.196	0.018	0.629	0.428	0.923
West Sources	0.411	0.159	0.01	1.508	1.104	2.06
East Sources	0.304	0.183	0.097	1.355	0.947	1.94
Water Source	0.177	0.694	0.799	1.193	0.306	4.649
Gender	−0.119	0.622	0.848	0.888	0.262	3.007
Age	0.031	0.031	0.316	1.032	0.971	1.097
Constant	−0.673					

Note: EM = emissions; B = unstandardized logit coefficient; S.E. = standard error; OR = odds ratio; CI = 95% confidence interval.

shows the annual median and upper and lower limits of ambient air levels of emissions of the group of compounds included in this analysis: nitrogen oxides, carbon monoxide, VOCs, PM 2.5, and formaldehyde. These five compounds had the highest reported mass and known health effects.

Relationships between exposure and respiratory health outcomes

Although all reported one or more symptoms that began or worsened after the onset of drilling activity and could not be plausibly attributed to pre-existing or current medical conditions, or practices such as smoking, 28% of the sample did not report any respiratory symptoms at all. At least one respiratory symptom (i.e., “any respiratory symptom”) was reported by 72%; sore throat by 40%; cough and shortness of breath by 36% each; sinus problems by 26%, and wheezing by 16%. The majority (77%) lived within 2 km of at least one source: 29% lived within 2 km of 1–9 sources; 25% within 2 km of 10–19 sources; and 23% within 2 km of 20 or more. The number of sources within 2 km ranged from 0 to 40.

Any respiratory symptom and exposure

For any respiratory symptom (i.e., at least one respiratory symptom reported), the overall model with 11 predictors was significant: $\chi^2(11) = 22.32$, $P = .022$ (Cox & Snell $r^2 = .231$; Nagelkerke $r^2 = .329$). As well, the Hosmer–Lemeshow test was not significant: $\chi^2(7) = 4.09$, $P = .769$. In regards to classification, sensitivity was 91.7% and specificity was 36%. As shown on Table 5, the following predictors are significant: (1) South Sources: $b = -.464$, $P = .018$ (OR = 0.629, 95% CI = 0.428, 0.923) indicating that the higher the value for south source the lower the probability of having any symptom; and (2) West Sources: $b = .41$, $P = .01$ (OR = 1.51, 95% CI = 1.10, 2.06) indicating that the higher the value for west source the higher the probability of having any symptom.

Table 6. Associations between exposure measures and cough.

	B	S.E.	*P*-value	OR	CI Lower	CI Upper
North EM	0	0	0.957	1	0.999	1.001
East EM	−0.001	0.001	0.084	0.999	0.998	1
South EM	0	0	0.4	1	1	1.001
West EM	−0.001	0	0.153	0.999	0.998	1
North Sources	−0.026	0.107	0.804	0.974	0.79	1.201
South Sources	−0.267	0.146	0.068	0.766	0.575	1.02
West Sources	0.209	0.102	0.04	1.232	1.009	1.505
East Sources	0.086	0.106	0.414	1.09	0.886	1.341
Water Source	2.218	0.684	0.001	9.186	2.403	35.125
Gender	0.404	0.618	0.514	1.497	0.446	5.029
Age	0.01	0.028	0.724	1.01	0.957	1.066
Constant	−2.528					

Note: EM = emissions; B = unstandardized logit coefficient; S.E. = standard error; OR = odds ratio; CI = 95% confidence interval.

Cough and exposure

For cough, the overall model with 11 predictors was significant: $\chi^2(11) = 27.06$, $P = .005$ (Cox & Snell $r^2 = .273$; Nagelkerke $r^2 = .373$). As well, the Hosmer–Lemeshow test was not significant: $\chi^2(7) = 8.03$, $P = .33$. In regards to classification, sensitivity was 67.7% and specificity was 74.1%. As shown in Table 6, the following predictors are significant: (1) West Sources: $b = .209$, $P = .04$ (OR = 1.23, 95% CI = 1.01, 1.51) indicating that the higher the value for west source the higher the probability of having the cough symptom and (2) water source: $b = 2.22$, $P = .001$ (OR = 9.19, 95% CI = 2.40, 35.13) indicating that those who have a well or other non- municipal water source have a higher probability of having the cough symptom than those who have a municipal water source.

Shortness of breath and exposure

For shortness of breath, the overall model with 11 predictors was significant: $\chi^2(11) = 25.54$, $P = .008$ (Cox & Snell $r^2 = .26$; Nagelkerke $r^2 = .355$). As well, the Hosmer–Lemeshow test was not significant: $\chi^2(7) = 1.73$, $P = .973$. In regards to classification, sensitivity was 45.2% and specificity was 87%. As shown on Table 7, the following predictors are significant: (1) South Sources: $b = -.372$, $P = .049$ (OR = 0.689, 95% CI = 0.476, 0.999) indicating that the higher the value for south source the lower the probability of having the shortness of breath symptom and (2) West Sources: $b = .439$, $P = .003$ (OR = 1.55, 95% CI = 1.16, 2.08) indicating that the higher the value for west source the higher the probability of having the shortness of breath symptom. There were no significant associations between the exposure measures and sore throat, sinus problems, or wheezing.

The results of our analysis suggest that an exposure metric including the number of sources in combination with wind direction may be a better predictor of new onset

Table 7. Associations between exposure measures and shortness of breath.

	B	S.E.	P-value	OR	CI Lower	CI Upper
North EM	0	0	0.546	1	1	1.001
East EM	−0.001	0	0.15	0.999	0.999	1
South EM	0.001	0.001	0.164	1.001	1	1.002
West EM	−0.002	0.001	0.078	0.998	0.996	1
North Sources	−0.142	0.124	0.254	0.868	0.68	1.107
South Sources	−0.372	0.189	0.049	0.689	0.476	0.999
West Sources	0.439	0.15	0.003	1.551	1.156	2.08
East Sources	0.259	0.156	0.097	1.295	0.954	1.759
Water Source	0.261	0.651	0.689	1.298	0.362	4.654
Gender	−1.027	0.586	0.08	0.358	0.114	1.129
Age	0.019	0.029	0.51	1.019	0.964	1.078
Constant	−1.361					

Note: EM = emissions; B = unstandardized logit coefficient; S.E. = standard error; OR = odds ratio; CI = 95% confidence interval.

respiratory symptoms than the number of sources and wind direction combined with the reported emissions from those sources.

There are several possible explanations for our findings. Any respiratory symptom, cough, and shortness of breath were associated with the cardinal direction of the emission source. The levels of contamination in the air move along with weather systems. In this county, that is primarily from the southwest to the northeast. The periods of stability in a weather system vary (i.e., there are periods of stability and low dilution of pollutants). It is possible that weather systems that move in other directions, although they also carry pollutants, have differing periods of high stability and low dilution.

Our calculations of the annual hourly emission rate from a source assume a consistent rate and may not be reflective of the occurrence of peak emissions. Peak emissions may be more important in precipitating acute respiratory symptoms than median emissions. Proximity to wells inherently captures peak events. In a similar vein, emission data do not capture exposures to flaring, impoundments or indoor off-gassing from contaminated water, all of which might result in peak exposures precipitating respiratory symptoms. Since these sources are typically co-located with wells, proximity to sources would better reflect these exposures.

Although four of the five compounds included in this analysis have recognized acute respiratory effects (NO_x, PM2.5, VOCs, and formaldehyde), there may be other emissions with potent respiratory effects that were not included in the analysis. Future analyses could be limited to those emissions with established respiratory actions. Additionally, our results support the potential exposure presented by ground water. Ground water contamination associated with the gas wells may contribute to respiratory symptoms such as cough through off gassing during indoor use.

Our study used a convenience sample whose self-reported date of respiratory symptom onset fell within the year of the exposure estimate. The lack of precision in the temporal relationship to the exposure is a limitation of this study. However, self-reported symptoms were reviewed by a nurse practitioner, in a standardized clinical interview, and all records were reviewed to include only those symptoms that could not plausibly be explained by a co-occurring medical condition or a habit such as smoking. A further limitation is our focus on respiratory symptoms, when there are other health symptoms that have been associated with UNGD that we did not include.

Other approaches to estimation of exposures have considered distance from nearest wells and the number of wells. Our approach considered the intensity of the emissions; temporal dilution factors; number of sources within a specified distance; and cardinal direction of those sources relative to the residence. The approach included not just well pads, but also processing plants and compressor stations. We feel that inclusion of infrastructure such as processing plants and compressor stations and the addition of cardinal direction provides a more precise assessment of exposure at the residence. The processes needed to consider cardinal directions are not overly cumbersome. When using approaches such as the inverse ratio of the distance squared, one can simply consider the direction from the source and flow of weather patterns in the area.

Conclusion

To our knowledge this is only the second study that included multiple sources of pollution (i.e., well pads, processing plants, and compressor stations) and the first study to incorporate weather and atmospheric conditions in the exposure estimate. Estimates of exposure typically characterize the sources within a specific radius of a residences. We suggest that future characterizations should consider the cardinal direction of the source from the residence. The impact of ambient air levels is unclear should be investigated in future studies.

Funding

This work was supported by funding from Heinz Endowments, Pittsburgh, PA.

ORCID

Leslie Walleigh http://orcid.org/0000-0003-1616-7676

References

[1] Schraufnagel, D. E.; Balmes, J. R.; Cowl, C. T.; De Matteis, S.; Jung, S.-H.; Mortimer, K.; Perez-Padilla, R.; Rice, M. B.; Riojas-Rodriquez, H.; Sood, A.; et al. Air Pollution and Noncommunicable Diseases: A Review by the Forum of International Respiratory Societies' Environmental Committee, Part 1: The Damaging Effects of Air Pollution. *Chest* **2019**, *155*, 409–416. DOI: 10.1016/j.chest.2018.10.042.

[2] Schraufnagel, D. E.; Balmes, J. R.; Cowl, C. T.; De Matteis, S.; Jung, S.-H.; Mortimer, K.; Perez-Padilla, R.; Rice, M. B.; Riojas-Rodriquez, H.; Sood, A.; et al. Air Pollution and Noncommunicable Diseases: A Review by the Forum of International Respiratory Societies' Environmental Committee, Part 2: Air Pollution and Organ Systems. *Chest* **2019**, *155*, 417–426. DOI: 10.1016/j.chest.2018.10.041.

[3] McKenzie, L. M.; Guo, R.; Witter, R. Z.; Savitz, D. A.; Newman, L. S.; Adgate, J. L. Birth Outcomes and Maternal Residential Proximity to Natural Gas Development in Rural Colorado.

Environ. Health Perspect. **2014**, *122*, 412–417. DOI: 10.1289/ehp.1306722.

[4] Stacy, S. L.; Brink, L. L.; Larkin, J. C.; Sadovsky, Y.; Goldstein, B. D.; Pitt, B. R.; Talbott, E. O. Perinatal Outcomes and Unconventional Natural Gas Operations in Southwest Pennsylvania. *PLoS One* **2015**, *10*, e0126425. DOI: 10.1371/journal.pone.0126425.

[5] Casey, J. A.; Savitz, D. A.; Rasmussen, S. G.; Ogburn, E. L.; Pollak, J.; Mercer, D. G.; Schwartz, B. Unconventional Natural Gas Development and Birth Outcomes in Pennsylvania, USA. *Epidemiology* **2016**, *27*, 163–172. DOI: 10.1097/EDE.0000000000000387.

[6] McKenzie, L. M.; Allshouse, W. B.; Byers, T. E.; Bedrick, E. J.; Serdar, B.; Adgate, J. L. Childhood Hematologic Cancer and Residential Proximity to Oil and Gas Development. *PloS One* **2017**, *12*, e0170423. DOI: 10.1371/journal.pone.0170423.

[7] Tustin, A. W.; Hirsch, A. G.; Rasmussen, S. G.; Casey, J. A.; Bandeen-Roche, K.; Schwartz, B. S. Associations between Unconventional Natural Gas Development and Nasal and Sinus, Migraine Headache, and Fatigue Symptoms in Pennsylvania. *Environ. Health Perspect.* **2017**, *125*, 189–197. DOI: 10.1289/EHP281.

[8] Graham, J.; Irving, J.; Tang, X.; Sellers, S.; Crisp, J.; Horwitz, D.; Muehlenbachs, L.; Krupnick, A.; Carey, D. Increased Traffic Accident Rates Associated with Shale Gas Drilling in Pennsylvania. *Accid. Anal. Prev.* **2015**, *74*, 203–209. DOI: 10.1016/j.aap.2014.11.003.

[9] Casey, J. A.; Wilcox, H. C.; Hirsch, A. G.; Pollak, J.; Schwartz, B. S. Associations of Unconventional Natural Gas Development with Depression Symptoms and Disordered Sleep in Pennsylvania. *Sci. Rep.* **2018**, *8*, 11375.

[10] Rasmussen, S. G.; Ogburn, E. L.; McCormack, M.; Casey, J. A.; Bandeen-Roche, K.; Mercer, D. G.; Schwartz, B. S. Association between Unconventional Natural Gas Development in the Marcellus Shale and Asthma Exacerbations. *JAMA Intern. Med.* **2016**, *176*, 1334–1343. DOI: 10.1001/jamainternmed.2016.2436.

[11] Rabinowitz, P. M.; Slizovskiy, I. B.; Lamers, V.; Trufan, S. J.; Holford, T. R.; Dziura, J. D.; Peduzzi, P. N.; Kane, M. J.; Reif, J. S.; Weiss, T. R.; Stowe, M. H. Proximity to Natural Gas Wells and Reported Health Status: Results of a Household Survey in Washington County, Pennsylvania. *Environ. Health Perspect.* **2015**, *123*, 21–26. DOI: 10.1289/ehp.1307732.

[12] Weinberger, B.; Greiner, L. H.; Walleigh, L.; Brown, D. Health Symptoms in Residents Living near Shale Gas Activity: A Retrospective Record Review from the Environmental Health Project. *Prev. Med. Rep.* **2017**, *8*, 112–115. DOI: 10.1016/j.pmedr.2017.09.002.

[13] Hill, E. L. Shale Gas Development and Infant Health: Evidence from Pennsylvania. *J Health Econ.* **2018**, *61*, 134–150. DOI: 10.1016/j.jhealeco.2018.07.004.

[14] Koehler, K.; Ellis, J. H.; Casey, J. A.; Manthos, D.; Bandeen-Roche, K.; Platt, R.; Schwartz, B. R. Exposure Assessment Using Secondary Data Sources in Unconventional Natural Gas Development and Health Studies. *Environ. Sci. Technol.* **2018**, *52*, 6061–6069. DOI: 10.1021/acs.est.8b00507.

[15] Pennsylvania Department of Environmental Protection. Air emissions inventory for the natural gas industry fact sheet [Internet]. [cited 2019 Mar 25]. Available from: http://files.dep.state.pa.us/Air/AirQuality/AQPortalFiles/Regulations%20and%20Clean%20Air%20Plans/Natural_Gas_Inventory_Fact_Sheet_02-11-13.pdf

[16] Air Emissions Report - Power BI Report Server [Internet]. [cited 2019 Mar 17]. Available from: http://www.depgreenport.state.pa.us/powerbiproxy/powerbi/Public/DEP/AQ/PBI/Air_Emissions_Report

[17] Brown, D. R.; Lewis, C.; Weinberger, B. I. Human Exposure to Unconventional Natural Gas Development: A Public Health Demonstration of Periodic High Exposure to Chemical Mixtures in Ambient Air. *J. Environ. Sci. Health Part A: Tox. Hazard Subst. Environ. Eng.* **2015**, *50*, 460–472. DOI: 10.1080/10934529.2015.992663.

[18] LatLong.net [Internet]. [cited 2019 Aug 3]. Available from: https://www.latlong.net/

[19] Pasquill, F. *Atmospheric Diffusion: The Dispersion of Windborne Material from Industrial and Other Sources*; D. Van Norstrand: London, **1962**.

[20] Burton, R. Atmospheric dispersion [Internet]. [cited 2019 Mar 24]. Available from: http://homepages.see.leeds.ac.uk/~lecrrb/dispersion/notes.pdf

[21] Brown, D.; Weinberger, B.; Lewis, C.; Bonaparte, H. Understanding Exposure from Natural Gas Drilling Puts Current Air Standards to the Test. *Rev. Environ. Health.* **2014**, *29*, 277–292. DOI: 10.1515/reveh-2014-0002.

[22] NOAA National Centers for Environmental Information. Data tools: Local climatological data [Internet]. [cited 2019 Mar 24]. Available from: https://www.ncdc.noaa.gov/cdo-web/datatools/lcd

[23] Menard, S. *Logistic Regression*; Sage: Los Angeles, CA, **2010**.

[24] Osborne, J. W. *Best Practices in Logistic Regression*; Sage: Los Angeles, CA, **2015**.

ATTACHMENT C

STUDY 21

Final Report

for

Pennsylvania Department of Health,

Bureau of Epidemiology

Hydraulic Fracturing Epidemiology Research Studies:

Asthma Outcomes

Prepared by:

University of Pittsburgh

School of Public Health

Contract number: 4400018535

July 31, 2023

University of Pittsburgh | School of Public Health

List of Contributors

Faculty
Jeanine M. Buchanich, MPH, MEd, PhD, Associate Professor, Department of Biostatistics
Principal Investigator, Retrospective Cohort Studies

Evelyn O. Talbott, MPH, DrPH, Professor of Epidemiology
Principal Investigator, Childhood Cancer Case-Control Study

Vincent Arena, PhD, Associate Professor Emeritus, Department of Biostatistics

Todd M. Bear, PhD, MPH, Assistant Professor, Department of Family Medicine, School of Medicine

James P. Fabisiak, PhD, Associate Professor, Department of Environmental and Occupational Health

Sally E. Wenzel, MD, Chair, Department of Environmental and Occupational Health

Ada O. Youk, PhD, Associate Professor, Department of Biostatistics

Jian-Min Yuan, MD, PhD, Professor, Department of Epidemiology

Staff
Jennifer Fedor, MS, Data Analysis

Allison C. Koller, MS, Project Coordinator

Michael F. Lann, MSIS, Programmer

David Maynard, Administrative Assistant

Andrew Mrkva, MA, Web Developer

Melanie Stangl, Administrative Assistant

Students
Jacob Cherry (Department of Biostatistics)
Violet Hecmanczuk (Department of Biostatistics)
Maggie Kuzemchak, MS (Department of Biostatistics)
Tianyuzhou (Jenny) Liang (Department of Biostatistics)
Emily Nicholls (Department of Environmental and Occupational Health)
Brian O'Connell, MS (Department of Biostatistics)
Nina Oryshkewych, MS (Department of Biostatistics)
Natalie Price (Department of Epidemiology)
Nicholas Tedesco (Department of Biostatistics)
Alex Watts (Department of Biostatistics)
Fan Wu, MPH (Department of Environmental and Occupational Health)

Consultant: Ravi Sharma, PhD, Adjunct Assistant Professor, Department of Epidemiology

Acknowledgments

The study team wishes to acknowledge and thank our External Advisory Board members for their thoughtful input throughout the study, including at meetings on August 5, 2021, November 10, 2021, January 26, 2022, April 13, 2022, and August 11, 2022.

External Advisory Board Members

David Allard, MS, Pennsylvania Department of Environmental Protection, Bureau of Radiation Protection (retired)

Uni Blake, PhD, American Petroleum Institute

Laura Dagley, Physicians for Social Responsibility (resigned)

Erica Jackson, FracTracker Alliance (resigned)

Ned Ketyer, MD, Physicians for Social Responsibility (resigned)

Jan Maund, (retired) Trinity Area School District

Curtis Schreffler, Deputy Director, Hydrologic Studies, PA Division, United States Geological Survey

Heaven Sensky, Center for Coalfield Justice (resigned)

Logan Spector, PhD, University of Minnesota Medical School

Judy Wendt Hess, PhD, Shell Oil

Edward Yorke, Canonsburg Councilman

We wish to thank the following organizations for providing study data. All interpretations are our own.

Pennsylvania Department of Health

Pennsylvania Department of Environmental Protection

University of Pittsburgh Biomedical Informatics Health Record Research Request (R3)

We also wish to thank Allison Hydzik, Media Relations, University of Pittsburgh Medical Center, for her invaluable assistance and advice.

Table of Contents

Tables

Figures

Asthma Outcomes Cohort Study

Background

Over the last 25 years, the American energy landscape has undergone an evolution, perhaps most notably with the expansion of hydraulic fracturing operations[1]. From 2000 to 2015, the number of hydraulically fractured wells in the United States increased from 23,000 to approximately 300,000. This rapid growth has corresponded to a range of economic benefits, including decreased energy costs and greatly increased production of both oil and natural gas[2]. However, mounting evidence suggests that hydraulic fracturing may have adverse impacts on public health and the environment[3-24].

Hydraulic fracturing – also known as fracking – is a process of unconventional natural gas development (UNGD) done by injecting large amounts of fluid at high pressure into dense rock in order to free trapped oil and natural gas[25]. The fluid used for injection typically consists of a mixture of water, sand (or other proppants), and various chemical additives. These wells, which are typically deeper than conventional wells, access previously unavailable reservoirs of oil and natural gas trapped in shale. The Marcellus Shale formation encompasses approximately half of Pennsylvania and is a large reservoir of natural gas.

Exposure to UNGD has been shown to be associated with some asthma exacerbations, including in Pennsylvania (PA)[8,13,26,27]. Rasmussen et al.[8] performed a case-control study using electronic health record data on 35,508 patients with asthma, aged 5 to 90 years, in Eastern PA. Patients with exacerbations from 2005 to 2012 were frequency matched on age, sex, and year of event to patients without an event. They assessed exposure to UNGD activity by well phase (well pad preparation, drilling, hydraulic fracturing, and production) the day before the event. Rasmussen et al. found statistically significant elevations in the highest tertile of activity, compared to the baseline of very low activity, for 11 of 12 UNGD phases by outcome models examined. The highest tertile odds ratios ranged from 1.45 for hospitalizations in the well pad preparation phase to 4.43 for mild exacerbations in the production phase.

Koehler et al.[13] (2018) used a principal components analysis to evaluate the association between mild asthma exacerbations, defined as new oral corticosteroid medications, and three UNGD activity metrics. They included nearly 70,000 exacerbations from approximately 40 counties in Eastern Pennsylvania. They constructed an exposure measure which included well pad development, drilling, stimulation, production, and compressor engines. Koehler et al. found statistically significantly elevated risk among those living within 1 kilometer (km) of the nearest well drilled, those in the highest tertile using an inverse distance weighted (IDW) cumulative metric, and those in the highest quartiles of exposure in the metrics that considered well phases plus distance.

There were three specific aims of this retrospective cohort study of asthma: 1) to replicate earlier studies conducted in Eastern PA using a population in Southwestern PA, where UNGD has proliferated in the past 15 years; 2) to enhance and improve upon previous UNGD exposure characterizations by assessing the associations between asthma exacerbations of various severity and each the four phases of UNGD; and 3) to enhance and improve upon previous UNGD exposure characterizations by assessing whether associations varied by multiple buffer distances to individuals' residences.

Methods

Asthma Records Data

Cohort members were identified from a University of Pittsburgh Health Record Research Request (R3) data request following University of Pittsburgh Institutional Review Board (IRB) approval. R3 is a service of the Department of Biomedical Informatics (DBMI) managed by the Chief Research Informatics Officer (CRIO), sponsored in part by the Clinical and Translational Sciences Institute and Institute for Precision Medicine at the University of Pittsburgh.

To be part of the cohort, participants need to have:

- An electronic health record with the University of Pittsburgh Medical Center (UPMC) health system between 2011-2020
- Age 5-90 years
- Patient residence within a zip code located within the eight-county study area (Allegheny excluding the City of Pittsburgh (excluded zip codes listed in Appendix Table 1), Armstrong, Beaver, Butler, Fayette, Greene, Washington, and Westmoreland counties)
- Primary diagnosis of asthma (codes shown in Appendix Table 2)
- At least one order for medications prescribed for asthma (Appendix Table 3)

We excluded participants with:

- Cystic fibrosis
- Pulmonary hypertension and pulmonary vascular disease (including pulmonary embolism)
- Paralysis of vocal cords or larynx
- Bronchiectasis
- Pneumoconiosis

ICD-9 and ICD-10 codes for inclusion and exclusion criteria are provided in Appendix Table 2.

Outcome Measures

Of interest were three levels of severity of asthma exacerbations among patients with asthma, defined according to the American Thoracic Society (ATS) and European Respiratory Society (ERS)[28]. Under the ATS/ERS criteria, patients with asthma are defined as those patients with a primary or secondary diagnosis of asthma (ICD-9 or ICD-10 codes; Appendix Table 2) in their electronic health record. Only exacerbation events among patients with at least one primary diagnosis were eligible for inclusion in the analysis. Exacerbation events were defined as follows:

1. **Severe exacerbation:** Initiation or increase of systemic corticosteroid medications among patients with asthma (Appendix Table 3).

2. **Emergency Department (ED) severe exacerbation:** ED or urgent care encounters for asthma that involve treatment with systemic corticosteroids among patients with asthma.

3. **Hospital exacerbation:** Hospitalizations for asthma that involve treatment with systemic corticosteroids among patients with asthma.

For patients with more than one type of exacerbation within 1 week, only the most severe exacerbation was retained. For patients with more than one exacerbation of a given type within a calendar year, one exacerbation of that type was randomly selected.

Control Selection

Controls were selected from patients in the study population. Patients with asthma who did not have an exacerbation during the study period were eligible to be controls for an exacerbation of any type. Patients with asthma who did have an exacerbation during the study period were eligible to be controls for a less severe exacerbation or an exacerbation of equal or greater severity up until the calendar year of their exacerbation.

Among eligible control patients, control events were a randomly selected contact date per calendar year per patient to replicate the methodology used in Rasmussen et al[8]. Contact dates were identified as all encounters with the health system recorded in the electronic health record (e.g., office visits, medication orders, procedures, tests, etc.).

Controls were frequency matched to cases by the following criteria: age category (5-12, 13-18, 19, 44, 45-61, 62-74, 75-90); sex (male, female); year of encounter.

Events

We restricted the pool of candidate case events to those among patients aged 5-90 years and living in a study area residence on the day of the event and the day prior. We randomly selected one residence for events associated in time with multiple residences (n=370). Finally, we randomly selected one event per type, per year, per patient to represent our final set of case events.

Control encounters were frequency matched to case events on patient age group, patient sex, and encounter year. We used 1:1 control: case matching for severe events, 2:1 matching for ED severe events, and 4:1 matching for hospitalization severe events.

Covariate Definitions

Clinical and demographic features of the patient and of the environment surrounding the patient's residence were included as covariates to control for potential confounding. Patient residences were extracted by R3 from the electronic health records and geocoded. Addresses for residences in rural zip codes were masked (a small amount of uncertainty was added by R3 to the latitude and longitude) prior to receipt of the data to avoid potential re-identification.

We received clinical and demographic covariates including patient sex, family history of asthma, and race/ethnicity. We also received clinical and demographic covariates that could change depending on the encounter, including age category, year of event, season of event, overweight and obesity status, smoking status, and Type II diabetes diagnosis. Event-level covariates included: year, season, age, BMI category, smoking status, average maximum temperature (degrees Celsius) recorded in the patient's county of residence on the day prior to the event, and community level socioeconomic deprivation index quartile. Covariate information is shown in Table 1.

Table 1. Covariates Included in the Analysis

Covariate	Definition
Patient level variables	
Patient sex	Male Female
History of asthma in patient's first-degree relatives (parents, siblings, offspring)	Yes No
Race (Self-reported race of the patient, categorized from 19 options)	White Black All other races Unknown
Ethnicity (Self-reported ethnicity of the patient)	Hispanic Not Hispanic Unknown
Variables that were dependent on the event date	
Event year	Calendar year in which the event occurred
Season in which the event occurred (based on month and day of the event)	Winter: December 22 – March 21 Spring: March 22 – June 21 Summer: June 22 – September 21 Fall: September 22 – December 21
Patient age category	Age in years at the time of the event, categorized as: 5-12, 13-18, 19-44, 45-61, 62-74, 75-90
Overweight and obesity status	Based on BMI calculated based on the weight in pounds and height in feet and inches at the event date or averaged across the visits before and after the event date[29] (Appendix Table 4).
Smoking status of the patient at the time of the event	Current Former Never Unknown
Type II diabetes diagnosis	Whether the patient had a diagnosis of type II diabetes (ICD-9 code 250.x0 and 250.x2 or ICD-10 code E11.x) at the time of the event (yes, no)
Maximum temperature on the previous day (°C)	Maximum recorded temperature in degrees Celsius on the date prior to the event date from the weather station nearest to each patient's residence. If data were missing for the nearest weather station, we used the county-level average maximum temperature.
Variables that were dependent on the event date and residence	
Community socioeconomic deprivation index	Quartiles (Q)1 – Q4 divided equally by the total number of communities in our study area

	Higher values of the index reflect greater community socioeconomic deprivation (Appendix Table 5 for details)

Exposure Measure

Unconventional natural gas development

The primary exposure measure was an inverse distance-weighted index of UNGD activity[6,8,11,13,15] up to 10 miles (or 16,093.4 m) of patient residence. Due to small numbers of asthma cases living within 0.5 miles of wells and the masking of rural geocodes performed by R3, we considered four buffer distances: 1 mile, 2 miles, 5 miles, and 10 miles in these models.

There are four phases of UNGD: well pad preparation, drilling, hydraulic fracturing, and production, which vary in duration and potential exposures. Information required to calculate the UNGD activity metric was obtained from the Pennsylvania Department of Environmental Protection (PA DEP) and Pennsylvania Department of Conservation and Natural Resources (PA DCNR).

1. **Well pad preparation -** the process of preparing a site where one or more wells were located. It is defined as the period 30 days before the first well on the pad is spudded.

2. **Drilling -** the creation of the wellbore. This phase begins on the well's spud date and ends on the drilling completion date.

3. **Hydraulic fracturing** (fracking, stimulation) **-** the process of injecting large volumes of water at high pressure into the wellbore to fracture the shale layer. This period is defined as beginning on the stimulation commencement date and ending on the stimulation completion date. Hydraulic fracturing may be repeated over time for a given well.

4. **Production -** the process of collecting natural gas or oil that, following hydraulic fracturing, travels through the wellbore to the surface. Production durations are variable; produced gas volume was represented as an average daily gas volume. A well was defined as being in production for reporting periods when production is indicated and reported production volume is non-zero.

Phase-specific UNGD metrics were calculated for each exacerbation using the following equations in Table 2.

Table 2. Definition of UNGD activity metric phase durations

Phase	Phase name	Calculation of phase-specific activity metric
1	Well pad preparation	Phase 1 metric for patient *j* event *k* = $\sum_{i=1}^{n} \frac{1}{d_{ijk}^2}$ Where: • *n* is the number of well pads in development within 10 miles of the residence of patient *j* on the day prior to event *k*

		• d^2_{ijk} is the squared distance (m^2) between well pad i and the residence of patient j at the time of event k
2	Drilling	Phase 2 metric for patient j event k = $\sum_{i=1}^{n} \frac{1}{d_{ijk}^2}$ Where: • n is the number of wells in the drilling phase within 10 miles of the residence of patient j on the day prior to event k • d^2_{ijk} is the squared distance (m^2) between well i and the residence of patient j at the time of event k
3	Hydraulic fracturing	Phase 3 metric for patient j event k = $\sum_{i=1}^{n} \frac{w_i}{d_{ijk}^2}$ Where: • n is the number of wells in the hydraulic fracturing phase within 10 miles of the residence of patient j on the day prior to event k • w_i is the depth (m) of well i • d^2_{ijk} is the squared distance (m^2) between well i and the residence of patient j at the time of event k
4	Production	Phase 4 metric for patient j event k =$\sum_{i=1}^{n} \frac{v_i}{d_{ijk}^2}$ Where: • n is the number of wells in production within 10 miles of the residence of patient j on the day prior to event k • v_i is the produced gas volume (m^3) of well i on the day prior to event k • d^2_{ijk} is the squared distance (m^2) between well i and the residence of patient j at the time of event k

Figure 1 illustrates the calculation of the phase-specific and buffer-specific metrics.

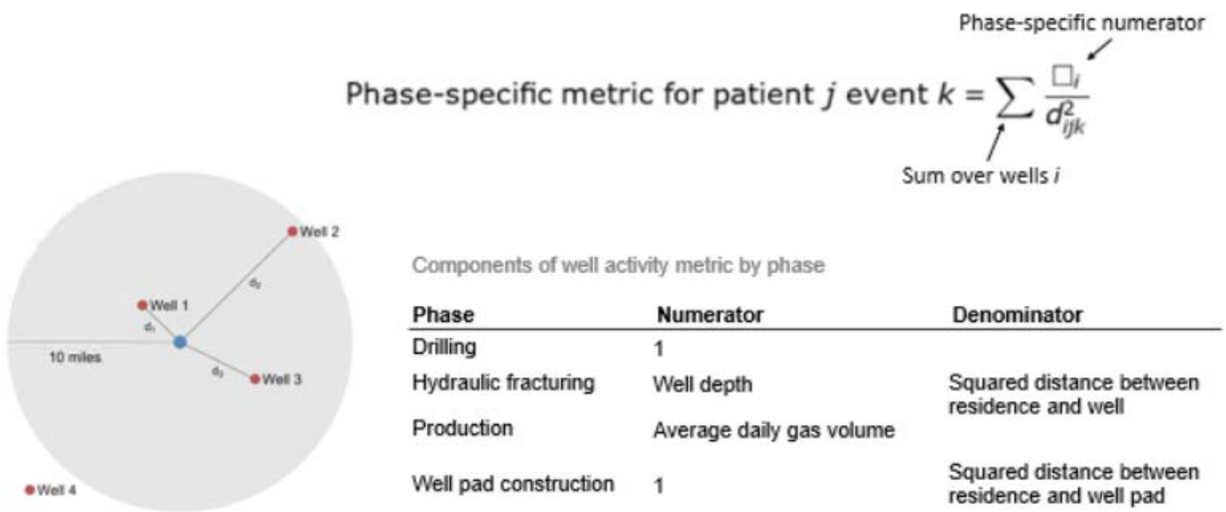


Figure 1. Well Phase Metric Calculation

We defined tertiles for each exposure metric (well pad construction phase, drilling phase, hydraulic fracturing phase, production phase) within each buffer distance (1, 2, 5, 10 miles):

- **Unexposed**: metric = 0
- **Exposed, low**: metric >0 and metric <33.3% of non-zero values among the entire cohort
- **Exposed, moderate**: metric >0 and metric ≥ 33.3% of non-zero values and metric <66.7% of non-zero values among the entire cohort
- **Exposed, high**: metric >0 and metric ≥ 66.7% of non-zero values among the entire cohort

Data Analysis

Data cleaning

We used graphical analyses, descriptive statistics, and exploratory data analysis to identify outlying observations, implausible values, and other inconsistencies, which were handled on a case-by-case basis, which occurred very infrequently. We examined all data for missingness. We computed the proportion of missing data for each variable contributing to the calculation of the exposure metric, the outcome variables, and the covariates. We stratified these calculations by year to examine patterns of missingness over time. We had no missing outcome information. If the proportions of missing covariate data were low (< 5%), we analyzed complete cases. We had greater than 5% unknown for BMI and smoking data. For BMI, we averaged BMI from the dates one year prior to the event date. Similarly, for smoking status, if a patient did not have a known smoking status on the event date, the most recent known smoking status prior to the event date was used.

For the UNGD exposure metric, we imputed missing well data using other available data. Missing well depths were imputed using the median well depth among wells not missing this measurement. Missing spud dates and stimulation dates were extrapolated using other available dates for each well and median phase durations among wells without missing dates.

Statistical analysis

We examined the four phases of the UNGD activity metric for correlation. In the event of substantial correlation among these four metrics, we would z-score each phase-specific metric, and then sum these z-scored phase-specific metrics to obtain a single, overall UNGD activity metric for each asthma exacerbation. However, we did not find evidence of correlation between the phases, and thus the phase-specific metrics were divided into three tertiles of exposure, representing low, moderate, and high UNGD activity, respectively.

We computed descriptive statistics (for continuous variables: mean and standard deviation or median and IQR; for categorical variables: frequency) for outcome variables and covariates. Descriptive statistics were calculated for each type of asthma exacerbation for cases and controls. We assessed differences in these distributions by running univariate logistic models using community as a random effect.

Our analyses assessed the association the phase-specific UNGD activity metrics (tertiles) with each of the three levels of asthma exacerbation severity. To do this, we fit a series of multilevel logistic regression models with a random intercept for community, as defined for the socioeconomic deprivation index, to account for nesting of patients within communities.

Each base model included all four phase-specific UNGD activity metrics. We then added to the base models: patient sex, year of encounter, race, family history of asthma, age category, smoking status, BMI category, season of event, type II diabetes diagnosis, community socioeconomic deprivation, and temperature (°C). We evaluated covariates for conditional significance (global tests assessing the covariate as a whole) using Wald or likelihood ratio tests . We also assessed trend for the tertiles of exposure using a Wald test for the linear form of the tertiles of exposure variables. We assessed multicollinearity among model covariates by calculating variance inflation factors (VIF).

Associations were reported as odds ratios comparing the tertile splits of the UNGD activity metric(s) to the unexposed group (reference level) with 95% confidence intervals. The odds ratio is used to determine whether a particular exposure (e.g., UNGD activity) is a risk factor for a particular type of asthma exacerbation, and to compare the magnitude of various risk factors for that outcome. Odds ratios (OR) can be interpreted as:

OR=1 Exposure (e.g., UNGD activity) does not affect odds of the type of asthma exacerbation

OR>1 Exposure (e.g., UNGD activity) is associated with higher odds of having the type of asthma exacerbation

OR<1 Exposure (e.g., UNGD activity) is associated with lower odds of having the type of asthma exacerbation

We used a two-sided type I error rate of 0.05 for significance testing. No adjustments were made for multiple comparisons. Analyses were performed using R version 4.1.2 (2021-11-01) and Stata 17 (StataCorp. 2021. *Stata Statistical Software: Release 17*. College Station, TX: StataCorp LLC). Forest plots were produced using GraphPad Prism version 9.5.1 for Windows (GraphPad Software, San Diego, California USA, www.graphpad.com). Forest plots are a graphical representation of odds ratios to facilitate comparisons across groups.

Results

Cohort Formation

Figure 3 shows the enrollment flowchart for the asthma cohort. We received 119,648 patients from R3, and our final cohort consisted of 46,676 patients.

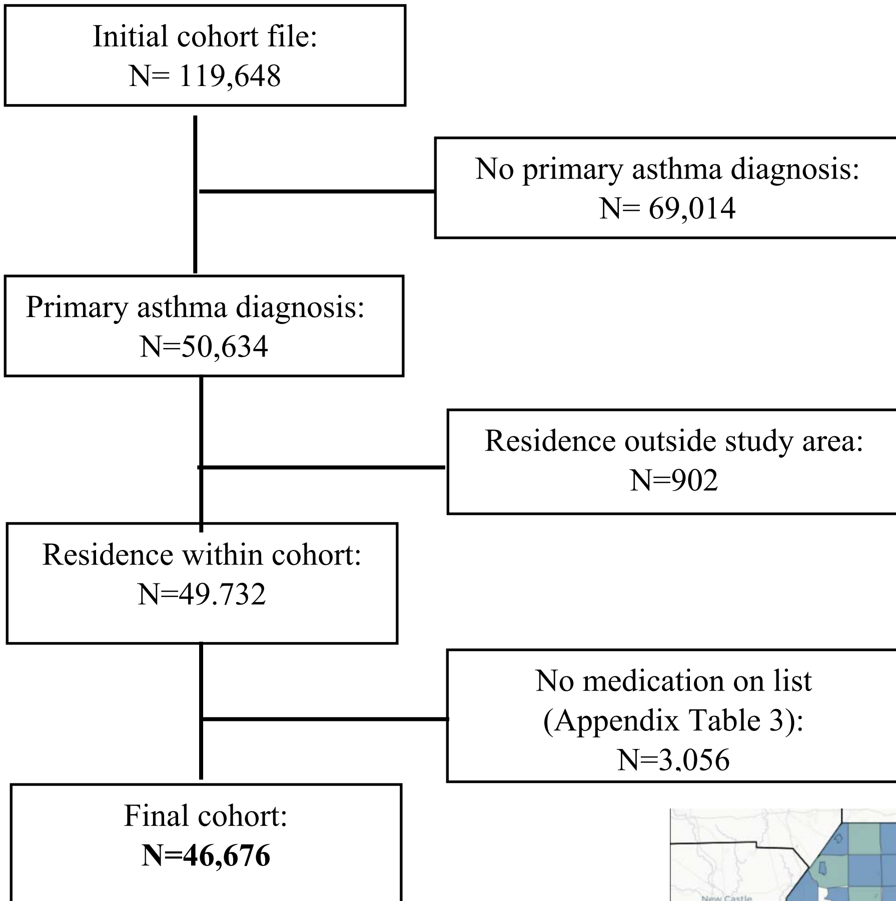


Figure 2. Cohort Enumeration Flowchart

The map (Figure 2) shows the counts for each patient community in the cohort. Allegheny County, excluding the City

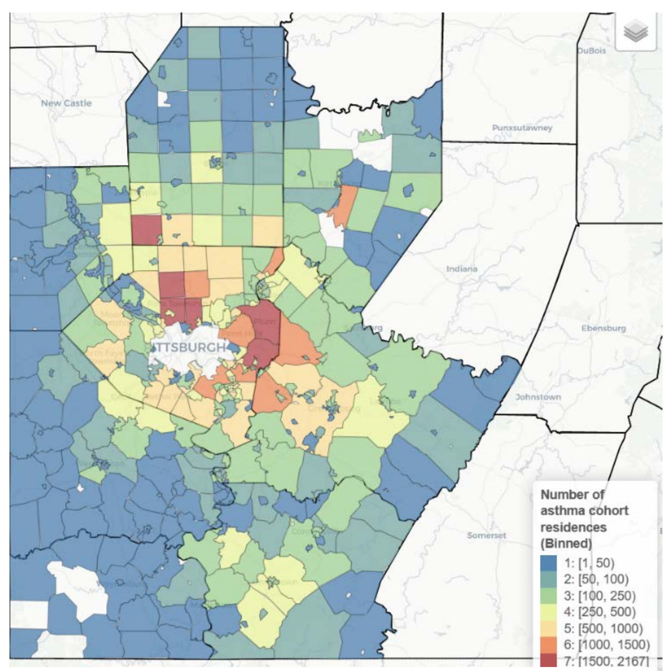


Figure 3. Map of Patient Communities

Events

One event per type, per year, per patient was randomly selected for our final set of case events. Table 3 shows the number of events per type and the number of controls. There were a total of 40,627 case and control events included.

Table 3. Final Counts by Event Type

Event Type	Number of Cases	Percent of Events	Number of Controls[1]
Severe	16,373	86.8	16,373
ED Severe	2,292	12.1	4,584
Hospitalization Severe	201	1.1	804

1- Control events frequency matched to case events by type: 1:1 severe, 2:1 ED severe; 4:1 hospitalization severe

Figure 4 shows the distribution of community socioeconomic deprivation index by quartile. Communities shown in blue are Quartile (Q) 1 (least deprivation) while communities shown in orange are in Q4 (most deprivation). Much of Allegheny County (excluding the City of Pittsburgh) and southern Butler County are in Q1.

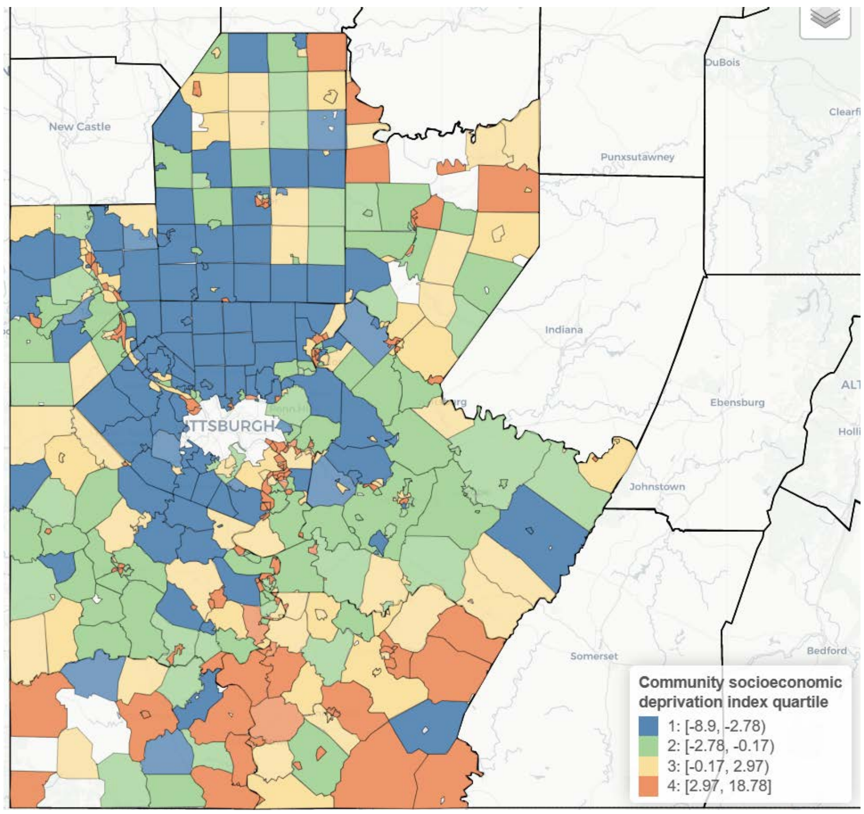


Figure 4. Map of Socioeconomic Deprivation Index by Community

UNGD Exposure

There were 5,799 wells included in our study from 2000 to 2020 (Figure 5). Through 2020, Washington County had the highest number of wells (n=1974), and Beaver County had the lowest number (n=141).

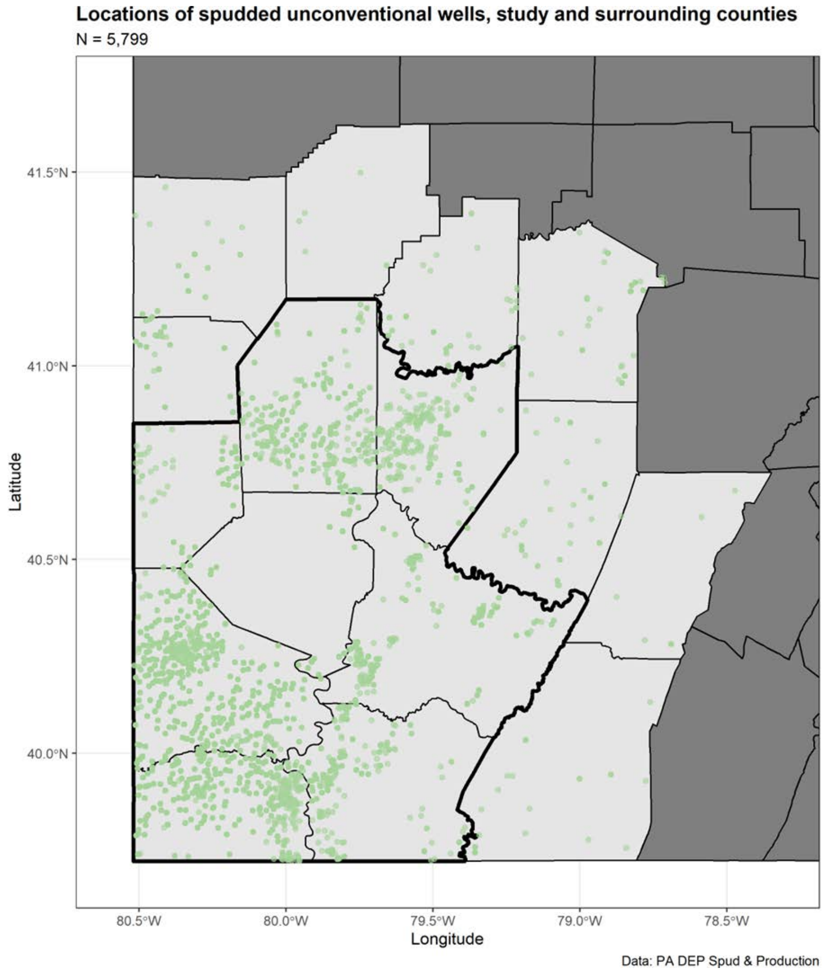


Figure 5. Map of UNGD Well Locations

There were fewer than 20 wells spudded in Southwestern Pennsylvania until 2007-2008, when production began increasing rapidly. The number of wells spudded peaked in 2014, with 765 as shown in Figure 6.

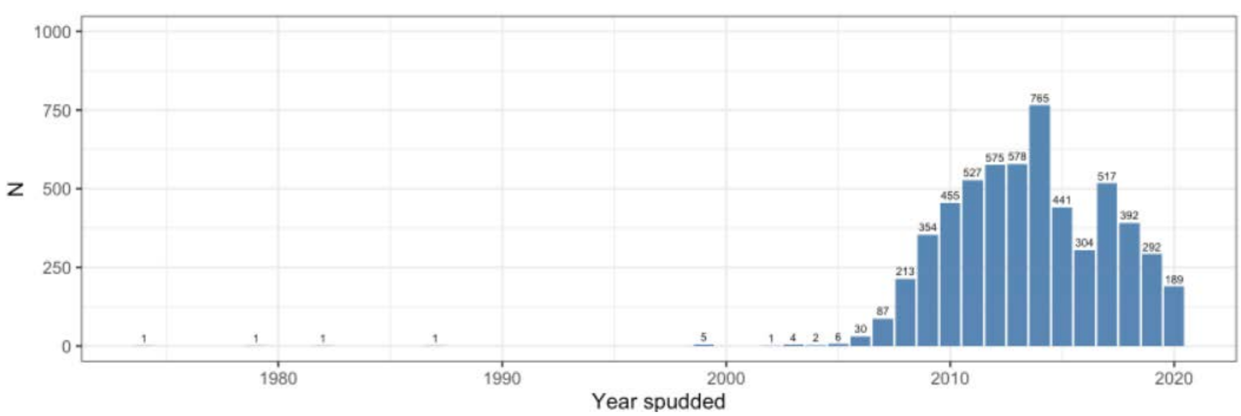


Figure 6. Histogram of UNGD Well Spud Dates by Year

Table 4 shows the median phase duration for each of the four UNGD activity metrics.

Table 4. UNGD Activity Metric Phase Durations

Phase	Phase name	Phase length
1	Well pad preparation	Minimum (spud date among wells on the pad) + 30 days 30 days
2	Drilling	Number of days between the spud and drilling completion dates Median: 104 days
3	Hydraulic fracturing	Number of days between stimulation commencement and stimulation completion Median: 12 days
4	Production	Duration of reporting period during which well reported production Mean: 2239 days (range 30-8769 days) Median: 2193 days

Shown in Table 5 are the cut points used for the tertiles (33.3% and 66.7%) as well as the minimum, median, and maximum value by phase and buffer. The production phase, which lasted the longest, had the highest metric values.

Table 5. Phase- and Buffer-Specific Cutpoints

Phase	Buffer (mi)	Min	33.3%	Median	66.7%	Max
Well pad preparation	1.0	3.88e-07	5.45e-07	7.96e-07	1.79e-06	7.31e-05
	2.0	9.70e-08	1.37e-07	1.75e-07	2.82e-07	7.31e-05
	5.0	1.54e-08	2.29e-08	3.02e-08	4.77e-08	7.32e-05
	10.0	3.86e-09	5.93e-09	8.38e-09	1.32e-08	7.32e-05
Drilling	1.0	3.94e-07	1.33e-06	2.05e-06	2.75e-06	3.09e-03
	2.0	9.65e-08	3.13e-07	5.02e-07	7.86e-07	3.09e-03
	5.0	1.54e-08	6.75e-08	1.08e-07	1.66e-07	3.09e-03
	10.0	3.86e-09	2.16e-08	3.56e-08	6.34e-08	3.09e-03
Hydraulic fracturing	1.0	1.38e-03	2.05e-03	2.32e-03	3.23e-03	2.15e-02
	2.0	2.10e-04	5.22e-04	7.08e-04	1.09e-03	2.15e-02
	5.0	3.05e-05	1.10e-04	1.72e-04	2.69e-04	2.15e-02
	10.0	6.47e-06	2.76e-05	4.23e-05	7.01e-05	2.15e-02
Production	1.0	1.65e-05	1.32e-02	4.46e-02	1.16e-01	3.73e+02
	2.0	3.69e-07	1.61e-02	3.30e-02	6.96e-02	3.73e+02
	5.0	7.23e-09	9.19e-03	2.40e-02	4.74e-02	3.73e+02
	10.0	6.62e-08	1.72e-02	1.72e-02	3.50e-02	3.73e+02

Characteristics by event type and case or control status are shown in Table 6, along with p-values assessing differences in distributions between the groups except for by sex, age or encounter year, which were matching variables. Severe exacerbations had statistically significant differences in distributions among cases and controls for all covariates. ED exacerbations had statistically significant differences in distributions among cases and controls for all covariates except family history of asthma and BMI. Hospitalizations had statistically significant differences in distributions among cases and controls in family history of asthma, BMI, and socioeconomic deprivation index.

Hospitalizations had the highest percentage of females. Severe exacerbations occurred most frequently among 5–13-year-olds, and ED and hospitalization exacerbations among 19–45-year-olds. Case events occurred most frequently in the winter and the majority of all patients were nonsmokers. More cases than controls were in Q1 (least deprived) and more controls than cases were in Q4 of the socioeconomic deprivation index for all event types.

Counts are also shown by event type by buffer distance, including for the 0.5-mile buffer. As shown, counts of exposed within the 0.5-mile buffer were so small as to preclude modeling. For all three event types, more case than control events were exposed for every well activity metric at every buffer distance for severe exacerbations. The greatest number of exposures occurred for the production phase among all events with up to 15% more case events exposed than control events.

Table 6. Descriptive Statistics of Cases and Controls by Asthma Exacerbation Type

	Severe Exacerbation		ED Exacerbation		Hospitalization Exacerbation	
Characteristic	Case, n=16,373 (%)	Control, n=16,373 (%)	Case, n=2292 (%)	Control, n=4584 (%)	Case, n=201 (%)	Control, n=804 (%)
Patient and Event Characteristics						
Female	9476 (57.9)	9476 (57.9)	1435 (62.6)	2870 (62.6)	141 (70.2)	564 (70.2)
Age in years, time of the event or matched encounter						
5 - <13	5065 (30.9)	5065 (30.9)	258 (11.3)	516 (11.2)	40 (19.9)	160 (19.9)
13 - <19	1710 (10.4)	1710 (10.4)	178 (7.8)	356 (7.8)	4 (2.0)	16 (2.0)
19 - <45	3425 (20.9)	3425 (20.9)	1048 (45.7)	2096 (45.7)	66 (32.8)	264 (32.8)
45 - <62	3533 (21.6)	3533 (21.6)	605 (26.4)	1210 (26.4)	52 (25.9)	208 (25.9)
62 - <75	2008 (12.3)	2008 (12.3)	172 (7.5)	344 (7.5)	29 (14.4)	116 (14.4)
75 - 90	632 (3.9)	632 (3.9)	31 (1.4)	62 (1.4)	10 (5.0)	40 (5.0)
Event year						
2011	1470 (9.0)	1470 (9.0)	93 (4.1)	186 (4.1)	11 (5.5)	44 (5.5)
2012	1625 (9.9)	1625 (9.9)	187 (8.2)	374 (8.2)	20 (9.9)	80 (9.9)
2013	1787 (10.9)	1787 (10.9)	216 (9.4)	432 (9.4)	20 (9.9)	80 (9.9)
2014	2168 (13.2)	2168 (13.2)	226 (9.9)	452 (9.9)	27 (13.4)	108 (13.4)
2015	2137 (13.1)	2137 (13.1)	234 (10.2)	468 (10.2)	26 (12.9)	104 (12.9)
2016	1640 (10.0)	1640 (10.0)	291 (12.7)	582 (12.7)	20 (9.9)	80 (9.9)
2017	1638 (10.0)	1638 (10.0)	257 (11.2)	514 (11.2)	22 (11.0)	88 (11.0)
2018	1502 (9.2)	1502 (9.2)	296 (12.9)	592 (12.9)	22 (11.0)	88 (11.0)
2019	1507 (9.2)	1507 (9.2)	324 (14.1)	648 (14.1)	18 (9.0)	72 (9.0)
2020	899 (5.5)	899 (5.5)	168 (7.3)	336 (7.3)	15 (7.5)	60 (7.5)
Family history of asthma (yes)	3044 (18.6)	2652 (16.2)	305 (13.3)	620 (13.5)	40 (19.9)	98 (12.2)
	<0.0001		0.772		0.007	

	Severe Exacerbation		ED Exacerbation		Hospitalization Exacerbation	
Characteristic	Case, n=16,373 (%)	Control, n=16,373 (%)	Case, n=2292 (%)	Control, n=4584 (%)	Case, n=201 (%)	Control, n=804 (%)
Race						
White	14,669 (89.6)	14,021 (85.6)	1881 (82.1)	3836 (83.7)	173 (86.1)	682 (84.8)
Black	1255 (7.7)	2, 011 (12.3)	202 (8.8)	661 (14.4)	24 (11.9)	105 (13.1)
Other/Unknown	448 (2.7)	351 (2.1)	209 (9.1)	87 (1.9)	4 (2.0)	17 (2.1)
	<0.0001		<0.0001		0.973	
Event season						
Winter: December 22 – March 21	4820 (29.4)	3884 (23.7)	688 (30.0)	1120 (24.4)	59 (29.4)	191 (23.8)
Spring: March 22 – June 21	3979 (24.3)	4045 (24.7)	533 (23.3)	1129 (24.6)	48 (23.9)	200 (24.9)
Summer: June 22 – September 21	2752 (16.8)	3990 (24.4)	402 (17.5)	1144 (25.0)	40 (19.9)	178 (22.1)
Fall: September 22 – December 21	4822 (29.5)	4454 (27.2)	669 (29.2)	1191 (26.0)	54 (26.9)	235 (29.2)
	<0.0001		<0.0001		0.416	
BMI						
Underweight or normal weight	5427 (33.1)	5744 (35.1)	636 (27.8)	1252 (27.3)	49 (24.4)	233 (29.0)
Overweight	3677 (22.5)	3559 (21.7)	613 (26.7)	1064 (23.2)	35 (17.4)	200 (24.9)
Obese	6852 (41.9)	6502 (39.7)	997 (43.5)	2141 (46.7)	111 (55.2)	351 (43.7)
Unknown	417 (2.5)	568 (3.5)	46 (2.0)	127 (2.8)	6 (3.0)	20 (2.5)
	<0.0001		0.126		0.014	
Smoking status						
Never	10,798 (65.9)	9922 (60.6)	1542 (67.3)	2518 (54.9)	122 (60.7)	450 (56.0)
Current	1466 (9.0)	1702 (10.4)	319 (13.9)	811 (17.7)	24 (11.9)	104 (12.9)
Former	2730 (16.7)	2827 (17.3)	396 (17.3)	957 (20.9)	43 (21.4)	180 (22.4)
Unknown	1379 (8.4)	1923 (11.7)	35 (1.5)	298 (6.5)	12 (6.0)	70 (8.7)
	<0.0001		<0.0001		0.480	
Community socioeconomic deprivation index, quartiles						
Q1	8875 (54.2)	8113 (49.6)	1315 (57.4)	2091 (45.6)	108 (53.7)	373 (46.4)
Q2	3365 (20.6)	3054 (18.7)	413 (18.0)	838 (18.3)	43 (21.4)	136 (16.9)
Q3	2083 (12.7)	2104 (12.8)	337 (14.7)	629 (13.7)	26 (12.9)	118 (14.7)
Q4	2050 (12.5)	3102 (18.9)	227 (9.9)	1026 (22.4)	24 (11.9)	177 (22.0)
	<0.0001		<0.0001		0.011	
Type II diabetes l (yes)	920 (5.6)	1599 (9.8)	78 (3.4)	515 (11.2)	25 (12.4)	78 (9.7)
	<0.0001		<0.0001		0.208	
Avg temperature day prior (°C) (SD)	14.8 (10.3)	16.8 (10.5)	15.1 (10.6)	16.8 (10.7)	15.7 (10.2)	16.2 (11.0)
	<0.0001		<0.0001		0.522	
Well Activity Metrics						
Exposed within 0.5-mile buffer						
Construction	8 (0.05)	7 (0.04)	1 (0.04)	1 (0.02)	0 (0.0)	0 (0.0)
Drilling	11 0.07)	3 (0.02)	4 (0.2)	1 (0.02)	0 (0.0)	0 (0.0)
Hydraulic fracturing	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Production	351 (2.1)	219 (1.3)	37 (1.6)	55 (1.2)	2 (1.0)	11 (0.01)
Exposed within 1 mile buffer						
Construction	24 (0.1)	18 (0.1)	4 (0.2)	1 (0.02)	1 (0.5)	0 (0.0)
Drilling	77 (0.5)	41 (0.3)	10 (0.4)	9 (0.2)	1 (0.5)	0 (0.0)
Hydraulic fracturing	9 (0.05)	5 (0.03)	1 (0.04)	1 (0.02)	0 (0.0)	0 (0.0)
Production	1131 (6.9)	842 (5.1)	119 (5.2)	211 (4.6)	11 (5.5)	30 (3.7)
Exposed within 2-mile buffer						
Construction	98 (0.6)	70 (0.4)	6 (0.3)	11 (0.2)	1 (0.5)	2 (0.2)
Drilling	369 (2.2)	262 (1.6)	42 (1.8)	55 (1.2)	3 (1.5)	11 (1.4)
Hydraulic fracturing	70 (0.4)	62 (0.4)	6 (0.3)	11 (0.2)	2 (1.0)	1 (0.1)

	Severe Exacerbation		ED Exacerbation		Hospitalization Exacerbation	
Characteristic	Case, n=16,373 (%)	Control, n=16,373 (%)	Case, n=2292 (%)	Control, n=4584 (%)	Case, n=201 (%)	Control, n=804 (%)
Production	3122 (19.1)	2270 (13.9)	368 (16.1)	637 (13.9)	32 (15.9)	107 (13.3)
Exposed within 5-mile buffer						
Construction	683 (4.2)	516 (3.1)	73 (3.2)	117 (2.6)	8 (4.0)	18 (2.2)
Drilling	2496 (15.2)	1883 (11.5)	302 (13.2)	512 (11.2)	28 (14.0)	74 (9.2)
Hydraulic fracturing	652 (4.0)	491 (3.0)	78 (3.4)	111 (2.4)	5 (2.5)	19 (2.4)
Production	8646 (52.8)	6853 (41.9)	1200 (52.3)	1890 (41.2)	99 (49.2)	317 (39.4)
Exposed within 10-mile buffer						
Construction	2706 (16.5)	2305 (14.1)	336 (14.7)	537 (11.7)	34 (16.9)	101 (12.6)
Drilling	7974 (48.7)	6937 (42.4)	1032 (45.0)	1841 (40.1)	89 (44.3)	312 (38.8)
Hydraulic fracturing	2937 (17.9)	2391 (14.6)	386 (16.8)	610 (13.3)	29 (14.4)	110 (13.7)
Production	14,825 (90.5)	13,133 (80.2)	2075 (90.5)	3686 (80.4)	183 (91.0)	625 (77.7)

Model Results

Severe Exacerbation Models

Adjusted models for severe asthma exacerbations are shown below. For the construction, drilling, and hydraulic fracturing phases, there were no consistent associations at any buffer distance. For the production phase, there were statistically significantly elevated odds ratios of 3 to 5 for all buffer distances, some of which increased with increasing intensity of exposure. For all buffer distances, both the global and trend p-values were statistically significant.

Table 7. Asthma Severe Exacerbation Models by Well Phase Activity Metric

Buffer	Adjusted OR[1] (95% CI)			
	Well Pad Preparation	Drilling	Hydraulic Fracturing	Production
1 mile				
Unexposed	--	--	--	--
Low	1.50 [0.53, 4.25]	1.81 [0.92, 3.54]	0.94 [0.15, 5.88]	3.80 [3.09, 4.67]**
Moderate	0.96 [0.29, 3.13]	1.36 [0.70, 2.65]	3.44 [0.37, 32.20]	3.83 [3.13, 4.67]**
High	0.65 [0.21, 1.95]	1.58 [0.77, 3.23]	0.87 [0.12, 6.30]	3.81 [3.11, 4.66]**
Global p-value	0.87	0.02	0.04	<0.0001
Trend p-value	0.07	<0.0001	<0.0001	<0.0001
2 miles				
Unexposed	--	--	--	--
Low	1.49 [0.84, 2.65]	1.01 [0.75, 1.36]	0.62 [0.35, 1.11]	4.52 [3.89, 5.25]**
Moderate	0.55 [0.32, 0.94]	1.22 [0.91, 1.63]	1.11 [0.59, 2.10]	5.12 [4.41, 5.95]**
High	1.11 [0.63, 1.96]	1.05 [0.79, 1.41]	0.98 [0.51, 1.90]	4.02 [3.45, 4.67]**
Global p-value	0.15	0.14	0.02	<0.0001
Trend p-value	0.09	<0.0001	0.001	<0.0001
5 miles				
Unexposed	--	--	--	--
Low	1.03 [0.83, 1.27]	1.12 [0.99, 1.26]	1.17 [0.94, 1.45]	4.41 [3.92, 4.96]**
Moderate	1.00 [0.82, 1.23]	1.10 [0.98, 1.24]	1.06 [0.86, 1.32]	4.63 [4.10, 5.24]**
High	0.93 [0.75, 1.14]	1.05 [0.93, 1.19]	0.99 [0.80, 1.22]	4.73 [4.14, 5.39]**
Global p-value	0.96	0.25	0.14	<0.0001
Trend p-value	0.36	0.01	0.22	<0.0001
10 miles				
Unexposed	--	--	--	--
Low	0.97 [0.87, 1.08]	1.02 [0.95, 1.10]	1.04 [0.93, 1.15]	3.53 [3.20, 3.89]**
Moderate	1.05 [0.95, 1.17]	1.07 [1.00, 1.15]	1.11 [1.00, 1.23]	4.29 [3.85, 4.78]**
High	0.99 [0.88, 1.10]	1.07 [0.99, 1.16]	1.10 [0.99, 1.22]	4.72 [4.18, 5.34]**
Global p-value	0.69	0.21	0.11	<0.0001
Trend p-value	0.52	0.12	0.10	<0.0001

1- Models adjusted for SES, encounter year, age category, sex, race, season, BMI category, smoking status, family history of asthma, temperature, and history of type II diabetes

* p<0.05; ** p<0.001

The severe exacerbation forest plots by buffer distance for each phase are shown in Figure 7. The vertical line at 1 represents a null relationship; dots below 1 indicate reduced risk and dots above 1 indicate increased risk.

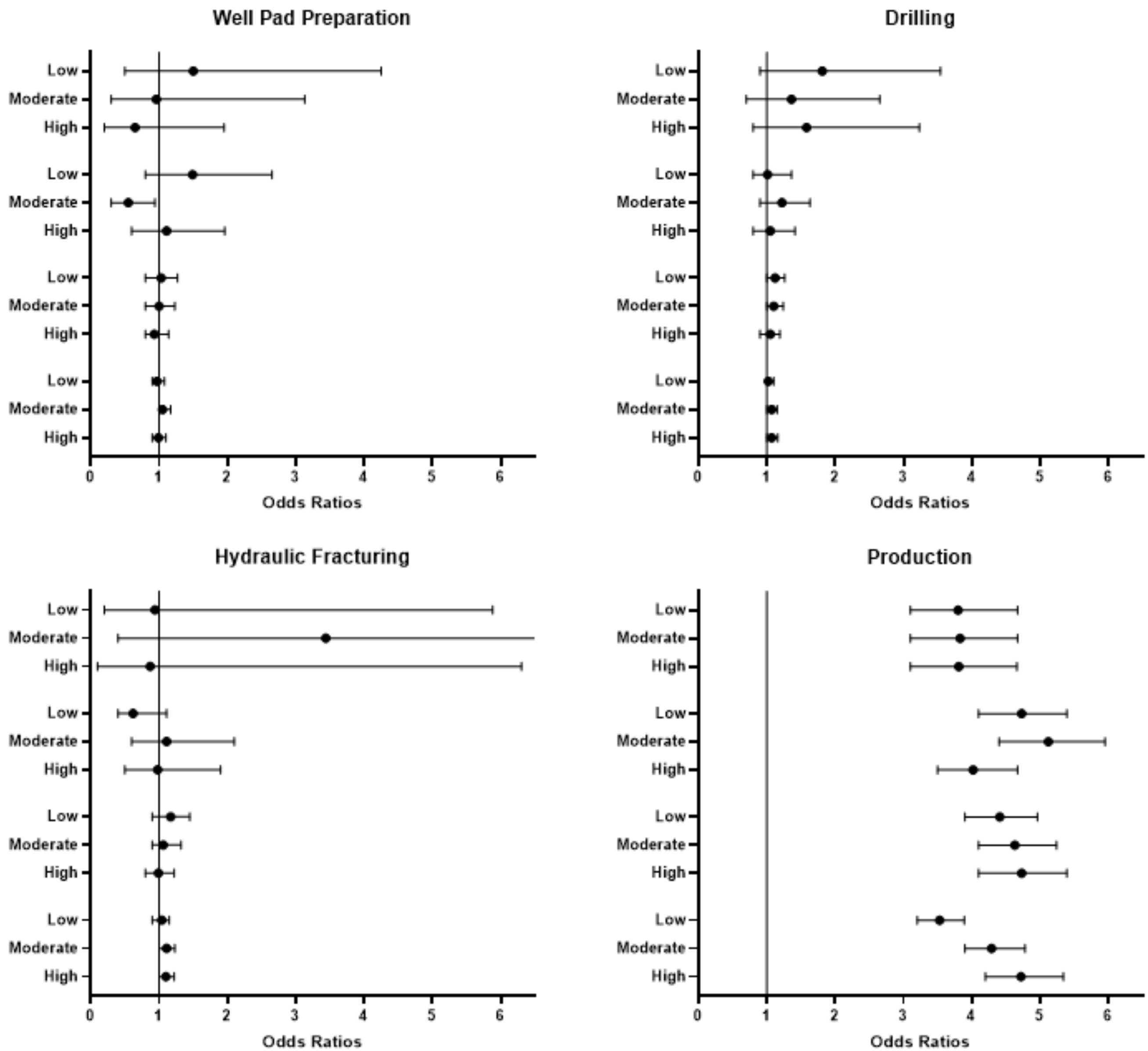


Figure 7. Forest Plots of Model Results for Severe Exacerbations

ED Severe Exacerbation Models

The adjusted models for exacerbations requiring an ED visit are shown in Table 8. Some of the exposure characterizations, noted as Not Applicable (NA), could not be modeled due to the number of cases and controls within the smaller buffer distances for this outcome (see Table 5 for counts of exposed cases and controls at each buffer distance). For the construction, drilling, and hydraulic fracturing phases, there were no consistent associations at any buffer distance. For the production phase, there were statistically significantly elevated odds ratios between 2 and 6 for all buffer distances, most of which increased with increasing intensity of exposure. For all buffer distances, both the global and trend p-values were statistically significant.

Table 8. Asthma ED Severe Exacerbation Model Results by Well Phase Activity Metric

Buffer	Adjusted OR[1] (95% CI)			
	Well Pad Preparation	Drilling	Hydraulic Fracturing	Production
1 mile				
Unexposed	NA[2]	NA	NA	--
Low				3.44 [1.85, 6.40]**
Moderate				3.96 [2.28, 6.87]**
High				4.86 [2.90, 8.16]**
Global p-value				<0.0001
Trend p-value				<0.0001
2 miles				
Unexposed	NA	--	NA	--
Low		0.98 [0.43, 2.24]		3.42 [2.33, 5.03]**
Moderate		1.63 [0.69, 3.85]		3.41 [2.37, 4.92]**
High		1.19 [0.49, 2.87]		4.13 [2.82, 6.05]**
Global p-value		0.74		<0.0001
Trend p-value		0.22		<0.0001
5 miles				
Unexposed	--	--	--	--
Low	0.87 [0.51, 1.50]	0.81 [0.60, 1.09]	0.98 [0.54, 1.79]	4.89 [3.65, 6.54]**
Moderate	1.55 [0.86, 2.81]	0.93 [0.67, 1.27]	1.17 [0.68, 2.01]	5.01 [3.71, 6.78]**
High	0.72 [0.36, 1.43]	0.90 [0.63, 1.27]	0.84 [0.41, 1.71]	4.11 [2.96, 5.70]**
Global p-value	0.31	0.71	0.77	<0.0001
Trend p-value	0.16	0.49	0.62	<0.0001
10 miles				

Buffer	Adjusted OR[1] (95% CI)			
	Well Pad Preparation	Drilling	Hydraulic Fracturing	Production
Unexposed	--	--	--	--
Low	0.96 [0.72, 1.28]	1.08 [0.91, 1.30]	0.93 [0.70, 1.24]	3.50 [2.75, 4.45]**
Moderate	1.32 [0.98, 1.77]	0.87 [0.72, 1.05]	1.23 [0.95, 1.61]	4.49 [3.45, 5.84]**
High	1.13 [0.83, 1.53]	0.82 [0.67, 1.01]	1.09 [0.82, 1.45]	4.81 [3.58, 6.47]**
Global p-value	0.28	0.07	0.39	<0.0001
Trend p-value	0.10	0.07	0.35	<0.0001

1- Models adjusted for SES, exposure year, age category, sex, race, season, BMI category, smoking status, family history of asthma, temperature, and history of type II diabetes

2- Small sample sizes precluded modeling

* p<0.05; ** p<0.001

Figure 8 shows the ED exacerbation forest plots by buffer distance for each phase. The vertical line at 1 represents a null relationship; dots below 1 indicate reduced risk and dots above 1 indicate increased risk.

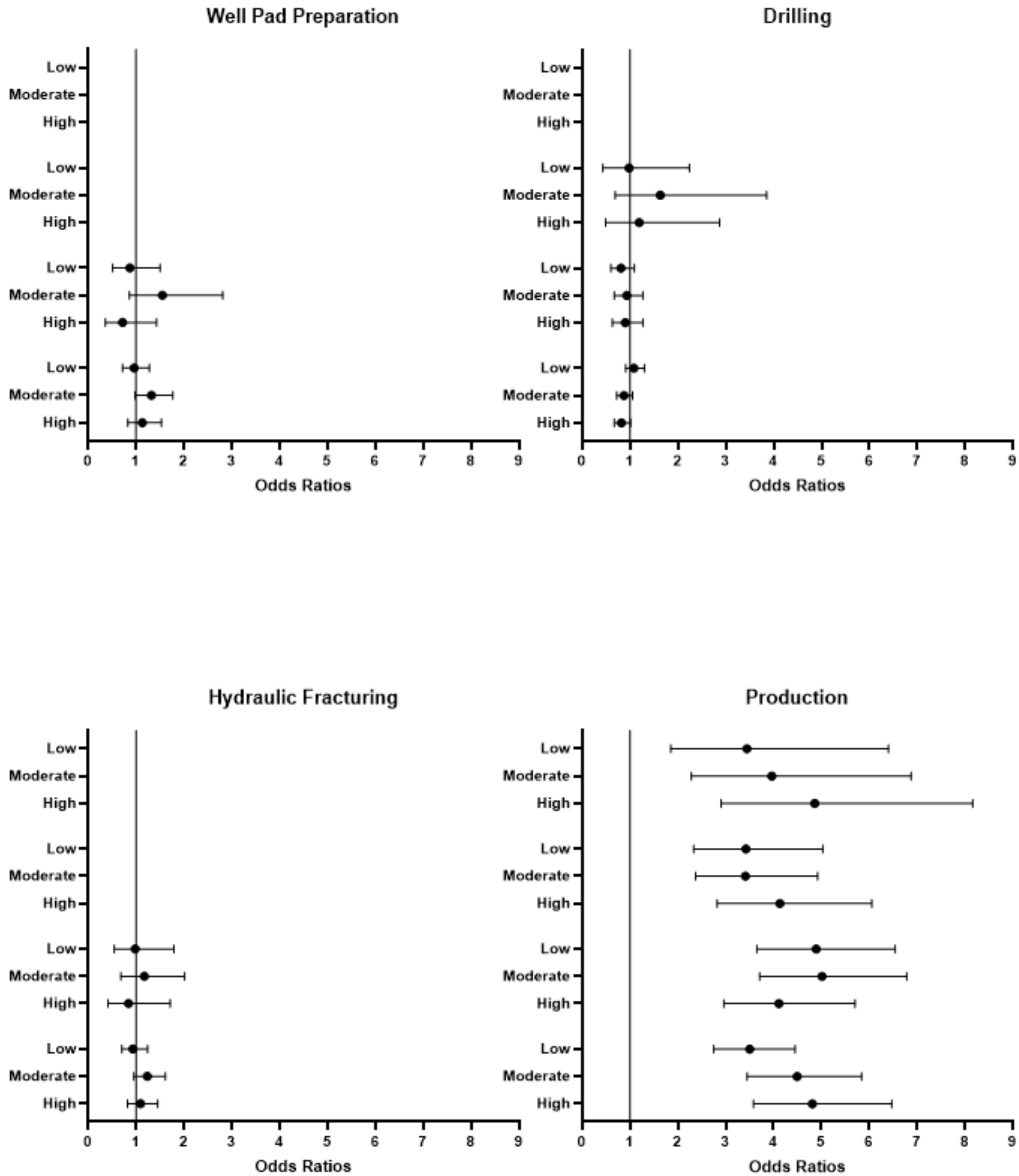


Figure 8. Forest Plots of Model Results for Emergency Department Severe

Hospitalization Severe Exacerbation Models
Adjusted model results for events requiring hospitalization are shown in Table 9. Some of the exposure characterizations, noted as Not Applicable (NA), could not be modeled due to the smaller number of cases (n=201) and controls (n=804) for this outcome (see Table 5 for counts of exposed cases and controls at each buffer distance). Only production could be modeled at the 1- and 2-mile buffers. For the construction, drilling, and hydraulic fracturing phases, there were no consistent associations at any buffer distance. For the production phase, all odds ratios were elevated and those odds ratios from 3 to 8 were statistically significantly. Most of the odds ratios increased with increasing intensity of exposure. For all buffer distances, both the global and trend p-values were statistically significant.

Table 9. Asthma Hospitalization Severe Exacerbation Model Results by Well Phase Activity Metric

Buffer	Adjusted OR[1] (95% CI)			
	Well Pad Preparation	Drilling	Hydraulic Fracturing	Production
1 mile				
Unexposed	NA[2]	NA	NA	--
Low				1.58 [0.30, 8.32]
Moderate				4.08 [1.01, 16.48]*
High				6.89 [1.54, 30.89]*
Global p-value				0.001
Trend p-value				0.001
2 miles				
Unexposed	NA	NA	NA	--
Low				2.01 [0.77, 5.26]
Moderate				2.33 [0.83, 6.55]
High				8.71 [3.09, 24.55]**
Global p-value				0.0001
Trend p-value				0.01
5 miles				
Unexposed	--	--	NA	--
Low	1.59 [0.35, 7.15]	0.67 [0.24, 1.86]		3.68 [1.79, 7.59]**
Moderate	0.93 [0.15, 5.79]	1.32 [0.52, 3.34]		3.08 [1.48, 6.42]*
High	1.75 [0.29, 10.57]	1.57 [0.66, 3.76]		4.77 [2.18, 10.45]**
Global p-value	0.95	0.58		0.0007
Trend p-value	0.52	0.40		0.01
10 miles				
Unexposed	--	--	--	--

Buffer	Adjusted OR[1] (95% CI)			
	Well Pad Preparation	Drilling	Hydraulic Fracturing	Production
Low	1.25 [0.61, 2.54]	0.85 [0.50, 1.45]	0.94 [0.44, 2.01]	3.13 [1.69, 5.81]**
Moderate	0.95 [0.42, 2.15]	0.64 [0.37, 1.13]	0.83 [0.37, 1.82]	3.64 [1.87, 7.09]**
High	0.83 [0.36, 1.87]	1.44 [0.81, 2.55]	0.49 [0.19, 1.26]	4.64 [2.25, 9.58]**
Global p-value	0.88	0.10	0.51	0.0003
Trend p-value	0.94	0.98	0.18	<0.0001

1- Models adjusted for SES, exposure year, age category, sex, race, season, BMI category, smoking status, family history of asthma, temperature, and history of type II diabetes

2- Small sample sizes precluded modeling

* p<0.05; ** p<0.001

The hospitalization severe exacerbation forest plots by buffer distance for each phase are shown in Figure 9. The vertical line at 1 represents a null relationship; dots below 1 indicate reduced risk and dots above 1 indicate increased risk.

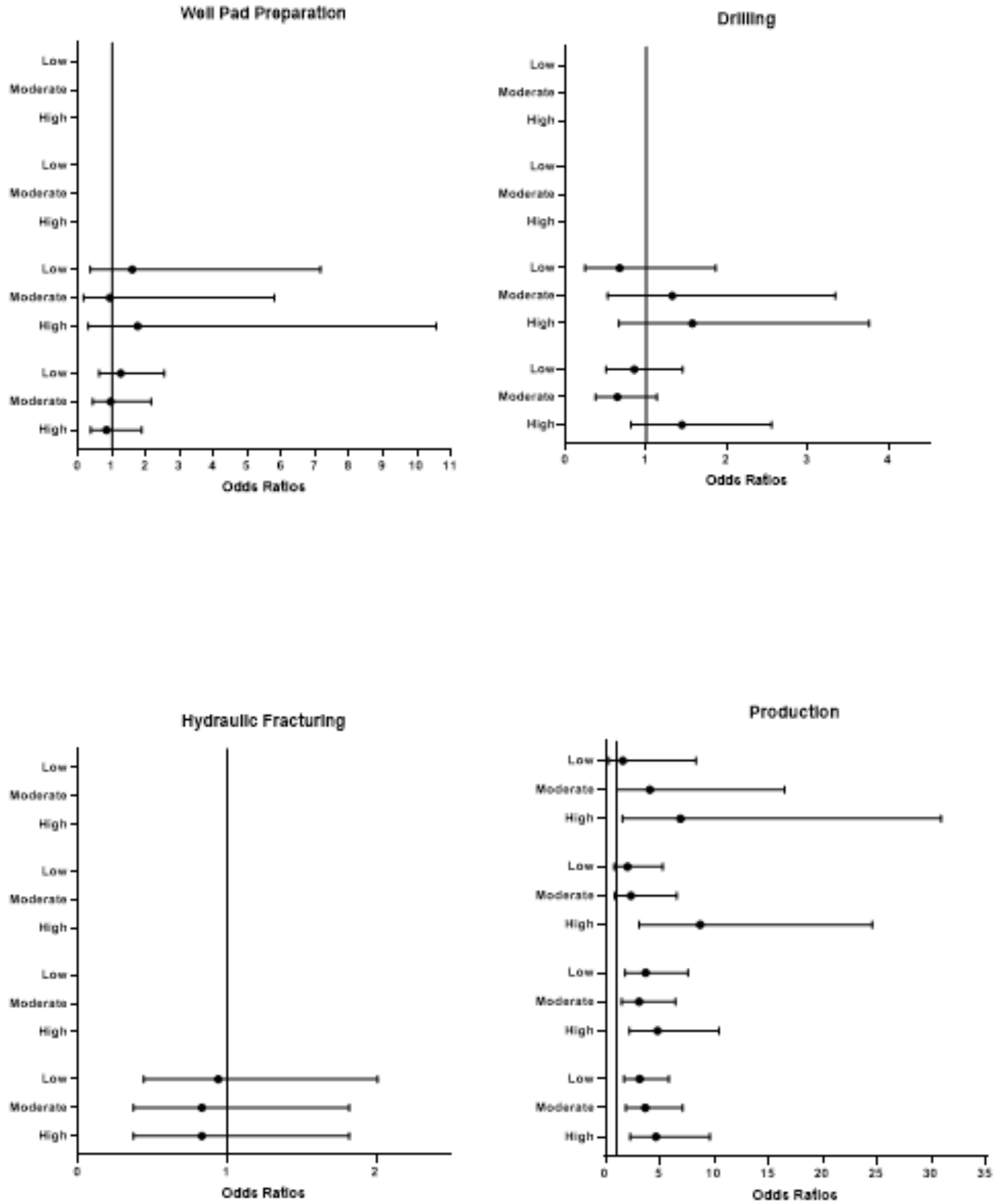


Figure 9. Forest Plots for Model Results for Hospitalization Severe Exacerbations

Discussion and Conclusions

This study examined three types of asthma events, severe, ED severe, and hospitalization severe, among more than 40,000 patients in an eight-county area of Southwestern PA from 2011-2020. To help frame the study conclusions, we are using the following classifying terms and criteria:

1. There are no data to suggest/support an increased risk:
   a. No statistically significantly elevated odds ratios
   b. Odds ratios at or near 1
   c. Odds ratios below 1 (with or without statistical significance)

2. There are limited data to suggest/support an increased risk:
   a. Statistically significantly elevated odds ratios in a low or moderate tertile
   b. Not statistically significant elevated odds ratios in multiple tertiles

3. There are moderate data to support an increased risk:
   a. Statistically significantly elevated odds ratios in multiple low or moderate tertiles
   b. Statistically significantly elevated odds ratios in a high tertile

4. There are strong data to suggest/support an increased risk:
   a. Statistically significantly elevated odds ratios in multiple tertiles
   b. Statistically significantly elevated odds ratios that increase across low, moderate, and high tertiles

We found strong evidence to suggest an increased risk in the production phase for all buffer distances examined for all three event types, based on consistent, statistically significantly elevated odds ratios. Elevations ranged from 2 to 8 times the baseline of no wells within 10 wells of the patient residence.

For all three event types, there were no data to support an increased risk at any buffer distance for the well pad preparation, drilling, and hydraulic fracturing phases.

This study replicated earlier work in Northeastern PA by Rasmussen et al[8]. In that study, they did not enforce buffer distances in their well activity metrics (all wells were included). Therefore, the most applicable comparison to these results is using our 10-mile buffer distances. Table 10 shows the odds ratios from this study compared to those from Rasmussen. The Rasmussen study found elevations in the well pad preparation, drilling, and hydraulic fracturing phases that were not found in this study.

Conversely, compared to Rasmussen, this study found much higher odds ratios for the production phase. Rasmussen reported an odds ratio of 4.43 in their highest tertile of production for their equivalent of our severe exacerbation, an order for an oral corticosteroid (OCS). That is similar to the odds ratio of 4.72 reported in this study. However, this study found elevated odds ratios for all tertiles of all buffers in the production phase.

Of note is that this study found the highest odds ratios for all asthma endpoints during the production phase. This could suggest that this phase may represent unique exposures not encountered during other phases. For example, this phase might be associated with more natural gas- and shale-derived hydrocarbons as well as produced water and, perhaps to some extent, hydraulic fracking fluid flowback. Moreover, this phase is generally the longest phase of well development and, thus, provides greater opportunity for chronic and cumulative exposures.

While this study had many similarities to that of Rasmussen et al, there are some notable differences. Rasmussen used a slightly less conservative definition for severe exacerbations of a new OCS medication ordered. We used the current ATS recommended definition of an initiation or increase of OCS medication in our three types of events. This could have led to less severe exacerbations included in the Rasmussen study than in ours.

Rasmussen et al. included events from 2005 to 2012 compared to 2011 to 2020 in our study. The 2011-to-2020-time frame was a particularly active time for UNGD development in Southwestern PA, but also encompassed technological advancements which may have modified exposure over time. We included encounter year as a covariate in our models to help account for these changes. Additionally, Rasmussen et al. did not enforce a buffer but included all wells in PA in their activity metrics, while this study specifically investigated the impacts at various buffer distances and excluded wells further than 10 miles from metric calculations.

Each study used electronic health records from a large provider in their region. The demographic characteristics are similar in Northeastern PA and our eight-county region in Southwestern PA, particularly with the exclusion of the City of Pittsburgh. However, Rasmussen et al. had a higher proportion of white patients in all case and control groups for all event types except ED cases. This study had a higher proportion of patients 5 to 13 years old than did Rasmussen et al. The high proportion of events among younger patients provides additional support that these are asthma exacerbations and not due to a chronic condition affecting older patients, such as chronic pulmonary obstructive disorder.

Strikingly, while our overall sample sizes were similar (n=46,676 patients in this study; n=35,508 patients in Rasmussen et al.), the Rasmussen et al. study had a much higher proportion of hospitalizations (n=4782 case events compared to n=201 case events in this study). This could be an indication of more poorly controlled asthma in that population which led to a higher proportion of very severe events. While we do not have similar information from Rasmussen, all of our hospitalization cases also had severe or ED exacerbations; there were no patients who only had a hospitalization during this timeframe. Among our severe exacerbations, only 7% (n=1122) had an ED or hospital exacerbation and 31.5% of our ED cases had severe (primarily) or hospitalization (rarely) exacerbations. This provides additional support that these findings are robust and are not being driven by a small number of patients with multiple endpoints.

Table 10. Comparison of Adjusted Odds Ratios in Current Study[1] with those in Rasmussen et al. (2016)[2]

	Severe Exacerbations		ED Exacerbations		Hospitalizations	
	Pitt SPH 10-mile buffer	Rasmussen 2016 (OCS[3] Orders)	Pitt SPH 10-mile buffer	Rasmussen 2016	Pitt SPH 10-mile buffer	Rasmussen 2016
Well Pad Preparation						
Unexposed	--	--	--	--	--	--
Low	0.97 (0.87, 1.08)	1.54 (1.37-1.74)*	0.87 (0.51, 1.50)	1.53 (1.06-2.23)*	1.25 (0.61, 2.54)	1.26 (1.06-1.50)*
Moderate	1.05 (0.95, 1.17)	1.66 (1.47-1.87)*	1.55 (0.86, 2.81)	1.77 (1.20-2.60)*	0.95 (0.42, 2.15)	1.37 (1.15-1.64)*
High	0.99 (0.88, 1.10)	1.59 (1.41-1.81)*	0.72 (0.36, 1.43	1.37 (0.94-1.99)	0.83 (0.36, 1.87)	1.45 (1.21-1.73)*
Drilling						
Unexposed	--	--		--	--	--
Low	1.02 (0.95, 1.10)	1.45 (1.29-1.63)*	0.81 (0.60, 1.09)	1.53 (1.06-2.21)*	0.85 (0.50, 1.45)	1.16 (0.98-1.37)
Moderate	1.07 (1.00, 1.15)	1.45 (1.29-1.63)*	0.93 (0.67, 1.27)	1.54 (1.04-2.27)*	0.64 (0.37, 1.13)	1.26 (1.05-1.50)*
High	1.07 (0.99, 1.16)	1.99 (1.75-2.26)*	0.90 (0.63, 1.27)	1.57 (1.08-2.29)*	1.44 (0.81, 2.55)	1.64 (1.38-1.97)*
Hydraulic Fracturing						
Unexposed	--	--		--	--	--
Low	1.04 (0.94, 1.15)	1.23 (1.09-1.39)*	0.93 (0.70, 1.24)	1.51 (1.05-2.19)*	0.94 (0.44, 2.01)	1.13 (0.96-1.33)
Moderate	1.11 (1.00, 1.23)	2.22 (1.95-2.53)*	1.23 (0.95, 1.61)	1.74 (1.17-2.61)*	0.83 (0.37, 1.82)	1.31 (1.10-1.57)*
High	1.10 (0.99, 1.22)	3.00 (2.60-3.45)*	1.09 (0.82, 1.45)	1.71 (1.16-2.52)*	0.83 (0.37, 1.82)	1.66 (1.38-1.98)*
Production						
Unexposed	--	--	--	--	--	--
Low	3.53 (3.20, 3.89)*	1.28 (1.13-1.46)*	3.50 (2.75, 4.45)*	1.47 (1.01-2.14)*	3.13 (1.69, 5.81)*	1.10 (0.92-1.30)
Moderate	4.29 (3.85, 4.78)*	2.15 (1.87-2.47)*	4.49 (3.45, 5.84)*	1.10 (0.74-1.65)	3.64 (1.87, 7.09)*	1.16 (0.97-1.38)
High	4.72 (4.18, 5.34)*	4.43 (3.75-5.22)*	4.81 (3.58, 6.47)*	2.19 (1.47-3.25)*	4.64 (2.25, 9.58)*	1.74(1.45-2.09)*

1 - Models adjusted for SES, exposure year, age category, sex, race, season, BMI category, smoking status, family history of asthma, temperature, and history of type II diabetes

2- From Rasmussen: Multilevel models with a random intercept for patient and community were adjusted for age category (5-12, 13-18, 19-44, 45-61, 62-74, >=75 years), sex (male or female), race/ethnicity (white, black, Hispanic, or other), family history of asthma (yes vs no), smoking status (never, former, current, or missing), season (spring, March 22–June 21; summer, June 22–September 21; fall, September 22–December 21; winter, December 22–March 21), Medical Assistance (yes vs no), overweight/obesity status (normal, body mass index [BMI], <85th percentile for children or <25 for adults; overweight, BMI, 85th to <95th percentile for children or 25 to <30 for adults; obese, BMI, >=95th percentile for children or>=30 for adults; or BMI missing), type 2 diabetes (yes vs no), community socioeconomic deprivation (across quartiles), distance to nearest major and minor arterial road (truncated at the 98th percentile, measured in meters, z transformed), squared distance to nearest major and minor arterial road (truncated at the 98th percentile, measured in meters, z transformed), maximum temperature on the day prior to event (measured in degrees Celsius), and squared maximum temperature on the day prior to event (measured in degrees Celsius).

3- Oral corticosteroids

* Statistically significant

Strengths and Limitations

This study has many strengths, including case ascertainment from a large health system with a large footprint in Southwestern Pennsylvania. However, we may have missed patients who used other health systems or facilities outside of this network for their care. We had few patients from Greene County; although this is the least populated county within our study area, it could also indicate that residents are receiving care outside of this network, including in neighboring West Virginia. We relied on electronic health records for our cohort information, which may not be reliable for some of our covariates, including but not limited to race and smoking status. These records may also fail to completely capture family history of asthma, and using ICD codes may not fully capture all cases of diabetes. This identification could be improved by including blood sugar and medication information. Additionally, individuals who do not have private insurance and those with more limited access to care could indicate a referral bias. This may partially explain the statistically significant differences among cases and controls for the socioeconomic deprivation index; there were fewer cases than controls for each event type in Quartile 4 (most deprivation).

The study applied a rigorous well phase activity assessment using multiple buffers to assess the strength of associations - the first to do so. These phase-by-buffer analyses provide new and important information about the associations of UNGD and asthma exacerbations. However, even in our large, system-based cohort, we had small sample sizes in some analyses, especially those within our smallest buffer distances and during shorter well activity phases (e.g., hydraulic fracturing). Additionally, we did not adjust for multiple comparisons. Some of the relationships between outcome and exposure may indicate evidence of a threshold effect, which was not assessed in the functional forms of the exposures examined here. Future studies should examine non-linear and other functional forms. The trend test assessed the linear relationship of the exposure tertiles, and some trend tests were statistically significant even when odds ratios (or term birthweights) were close to the reference level. Our well phase activity metric does not directly assess exposures to specific hazards associated with UNGD activity. The drop in cases in 2020 may indicate that we did not have complete coverage in that year but could also be an impact of the Covid-19 pandemic. However, our more than 10-year examination of cases lends additional credibility to these results. Additionally, the geocoding restrictions may have impacted exposure assignments at small buffers; however, we do not anticipate that this non-differential misclassification would have influenced the results.

Future analyses should consider a more direct exposure pathway than our UNGD metric. These results should also be examined by age group to understand whether those most vulnerable, including children and the elderly, are more strongly impacted. Additionally, we considered exposures only one day prior to the event. Other windows, including those from 2-5 days prior, should be examined to ensure the effects are similar.

Our UNGD exposure metric was based on residence in the electronic health records. However, exposures occur outside of the home as well, including at daycare, school, and work. Future work should consider the impact of these non-residential exposures as well. Additionally, as our buffer distances increased, the opportunity for non-well exposures increased. Asthma exacerbations could be associated with other additional exposures that may influence air quality, such as UNGD infrastructure and non-UNGD exposures.

Appendix

Table A1. List of zip codes located all or in part in the City of Pittsburgh in Allegheny County

Zip code	All or part City of Pittsburgh
15106	Part City
15120	Part City
15201	All City
15203	All City
15204	Part City
15205	Part City
15206	All City
15207	All City
15208	All City
15210	Part City
15211	All City
15212	Part City
15213	All City
15214	Part City
15215	Part City
15216	Part City
15217	All City
15218	Part City
15219	All City
15220	Part City
15221	Part City
15222	All City
15224	All City
15226	Part City
15227	Part City
15230	All City
15232	All City
15233	All City
15234	Part City
15235	Part City
15240	Part City
15260	All City
15282	All City

Table A2. **Study population inclusion and exclusion criteria ICD-9 and ICD-10 codes**

Name	**ICD-9 codes**	**ICD-10 codes**
Inclusion criteria		
Asthma	493.00, 493.01, 493.02, 493.10, 493.11, 493.12, 493.20, 493.21, 493.22, 493.81, 493.82, 493.90, 493.91, 493.92	J45.20, J45.22, J45.21, J45.990, J45.991, J45.909, J45.998, J45.902, J45.901
Exclusion criteria		
Cystic fibrosis	277.00, 277.01, 277.02, 277.03, 277.09	E84.9, E84.11, E84.0, E84.19, E84.8
Chronic pulmonary heart disease	416.0, 416.1, 416.2, 416.8, 416.9	I27.0, I27.1, 127.82, I27.2, I27.89, I27.81, I27.9
Paralysis of vocal cords or larynx	478.30, 478.31, 478.32, 478.33, 478.34	J38.00, J38.01, J38.02
Bronchiectasis	494.0, 494.1	J47.9, J47.1
Pneumoconiosis	500, 501, 502, 503, 504, 505, 506.0, 506.1, 506.2, 506.3, 506.4, 506.9, 507.0, 507.1, 507.8, 508.0, 508.1, 508.2, 508.8, 508.9	J60, J61, J62.8, J63.0, J63.1, J63.2, J63.3, J63.4, J63.5, J63.6, J66.0, J66.1, J66.2, J66.8, J64, J68.0, J68.1, J68.2, J68.3, J68.4, J68.9, J69.0, J69.1, J69.8, J70.0, J70.1, J70.5, J70.8, J70.9

Table A3. Oral corticosteroid medication order exclusion criteria ICD-9 and ICD-10 codes

Name	ICD-9 codes	ICD-10 codes
Suppurative and unspecified otitis media	382.00, 382.01, 382.02, 382.1, 382.2, 382.3, 382.4, 382.9	H66.009, H66.019, H67.9, H66.13, H66.23, H66.3X9, H66.40, H66.90
Non-suppurative otitis media and Eustachian tube disorders	381.00, 381.01, 381.02, 381.03, 381.04, 381.05, 381.06, 381.10, 381.19, 381.20, 381.29, 381.3, 381.4, 381.50, 381.51, 381.52, 381.60, 381.61, 381.62, 381.63, 381.7, 381.81, 381.89, 381.9	H65.199, H65.00, H65.119, H65.20, H65.30, H65.499, H65.90, H68.009, H68.019, H68.029, H68.109, H68.119, H68.129, H68.139, H69.00, H69.80, H69.90
Contact dermatitis and other eczema	692.0, 692.1, 692.2, 692.3, 692.4, 692.5, 692.6, 692.70, 692.71, 692.72, 692.73, 692.74, 692.75, 692.76, 692.77, 692.79, 692.81, 692.82, 692.83, 692.84, 692.89, 692.9	L24.0, L24.1, L24.2, L25.1, L25.3, L25.4, L25.5, L57.8, L55.0, L55.9, L56.0, L56.1, L56.2, L57.1, L57.5, L57.9, L56.5, L55.1, L55.2, L56.8, L25.0, L58.9, L23.0, L24.81, L23.81, L25.2, L25.8, L25.9
Other and unspecified disorders of back	724.00, 724.01, 724.02, 724.03, 724.09, 724.1, 724.2, 724.3, 724.4, 724.5, 724.6, 724.70, 724.71, 724.79, 724.8, 724.9	M48.00, M48.04, M48.06, M48.08, M54.6, M54.5, M54.30, M54.14, M54.15, M54.16, M54.17, M54.89, M54.9, M43.27, M43.28, M53.2X7, M53.3, M53.2X8, M54.08, M43.8X9, M53.9

Table A4. BMI cutoff values

For those aged 20 years or younger, we used the following criteria based on the CDC's recommended youth BMI-for-age cutoffs:
• **Underweight:** <5th percentile
• **Normal:** 5th to <85th percentile
• **Overweight:** 85th to <95th percentile
• **Obese:** ≥ 95th percentile
• Unknown: missing height and/or weight
For those aged 21 years or older, or when age was missing, we used the following criteria based on the CDC's recommended cutoffs for adults:
• **Underweight:** BMI <18.5
• **Normal:** BMI $\in$ [18.5, 25)
• **Overweight:** BMI $\in$ [25, 30)
• **Obese:** BMI ≥ 30
• Unknown: missing height and/or weight

Table A5. Calculation of Community Socioeconomic Deprivation Index

An index of socioeconomic deprivation incorporating six indicators from the 2015-2019 American Community Survey 5-year estimates from the US Census:
• Percent less than high school education
• Percent in poverty
• Percent not in the labor force
• Percent on public assistance
• Percent does not own a vehicle
• Percent civilian unemployment
The six indicators were standardized for direction, natural log-transformed, if necessary, z-scored using the standard deviations for Pennsylvania, and summed to create the final, unitless index for each county, township, or census tract. The total number of communities was divided into quartiles of socioeconomic deprivation index. Higher values of the index reflect greater community socioeconomic deprivation.

B. Detailed Cohort Characteristics

Patient level

Patient-level covariates included: race, ethnicity, sex, family history of asthma, and diagnosis of type II diabetes. Race, sex, and family history of asthma were time-invariant. Diagnoses of type II diabetes was time-varying in that they did not have the condition prior to their first diagnosis. Counts shown below for these variables are based on the total cohort of n=46,676 patients.

Race was self-reported in the EHR and was categorized from 19 options to the following, shown in the table below. Approximately 85% of patients identified as white and 12% identified as black.

Patients by collapsed race category

Race category	Number	Percent
White	39,621	84.9
Black	5,524	11.8
Unknown	894	1.9
All other races	637	1.4

Ethnicity was self-reported in the EHR and categorized as shown in the table below. Due to the very small proportion of Hispanic patients, this covariate was not included in the models.

Patients by collapsed ethnicity category

Ethnicity category	Number	Percent
Not Hispanic	44,414	95.2
Unknown	1,887	4.0
Hispanic	375	0.8

Self-reported sex was available from EHR. Nearly 60% of patients were female.

Patients by sex provided in EHR

Sex	Number	Percent
Female	27,337	58.6
Male	19,339	41.4

There were n = 7,209 patients (15.4%) with a family history of asthma. Most had a history of asthma in their biological mother only or biological father only.

Patients by family history of asthma including offspring as first-degree relatives

Family history of asthma	Number	Percent
No	39,467	84.6
Yes	7,209	15.4

About 8% of the cohort has at least one primary diagnosis for type II diabetes.

Patients by primary type II diabetes diagnosis

Type II diabetes	Number	Percent
No	42,865	91.8
Yes	3,811	8.2

Event-level

Visits by year are shown below. There were a higher proportion of visits in 2014 and 2015 and a lower proportion in 2020, which could be indicative of incomplete ascertainment for that year.

Year	Number	Percent
2011	3,274	8.06
2012	3,911	9.63
2013	4,324	10.64
2014	5,149	12.67
2015	5,106	12.57
2016	4,253	10.47
2017	4,158	10.23
2018	4,002	9.85
2019	4,076	10.03
2020	2,377	5.85

Season is shown below. Summer had the fewest number of events.

Season	Number	Percent
Fall	11,425	28.12
Spring	9,937	24.46
Summer	8,506	20.94
Winter	10,762	26.49

The frequency and percent by age group are shown below. Ages 5-13 had the greatest number of events while ages 75-90 had the fewest number.

Age Group	Number	Percent
[5, 13)	11,105	27.33
[13, 19)	3,974	9.78
[19, 45)	10,326	25.41
[45, 62)	9,141	22.50
[62, 75)	4,677	11.51
[75, 90]	1,407	3.46

Information on BMI is shown below. Less than 3% were missing BMI; 42% were obese and 33% were not overweight or obese.

BMI Category	Number	Percent
Not Overweight or Obese	13,341	32.84
Overweight	9,148	22.52
Obese	16,957	41.74
Missing	1,184	2.91

Smoking status is shown in the table below. The majority of events were associated with never smokers, while 9% of events had missing smoking information.

Smoking Status	Number	Percent
Never smoker	25,353	62.40
Current smoker	4,426	10.89
Former smoker	7,134	17.56
Unknown/missing data	3,717	9.15

There were 509 communities represented among the participants. The communities were divided into quartiles to form the cut points (approximately 127 communities in each quartile). Community-level socioeconomic deprivation index by quartile is shown below. Over half of the events were in communities in the highest (best) quartile; 16% were in the lowest (4th quartile).

SES Quartile	Number	Percent
Q1	20,875	51.38
Q2	7,851	19.32
Q3	5,297	13.04
Q4	6,607	16.26

References

1. Gallegos TJ, Varela BA. Trends in hydraulic fracturing distributions and treatment fluids, additives, proppants, and water volumes applied to wells drilled in the United States from 1947 through 2010: data analysis and comparison to the literature. Scientific Investigations Report. 2014-5131. Reston, VA: 2015. (http://pubs.er.usgs.gov/publication/sir20145131).
2. Clark C, Burnham A, Harto C, Horner R. Hydraulic fracturing and shale gas production: technology, impacts, and policy. 2012.
3. McKenzie LM, Witter RZ, Newman LS, Adgate JL. Human health risk assessment of air emissions from development of unconventional natural gas resources. Science of The Total Environment 2012;424:79-87. DOI: https://doi.org/10.1016/j.scitotenv.2012.02.018.
4. Adgate JL, Goldstein BD, McKenzie LM. Potential Public Health Hazards, Exposures and Health Effects from Unconventional Natural Gas Development. Environmental Science & Technology 2014;48(15):8307-8320. DOI: 10.1021/es404621d.
5. Stacy SL, Brink LL, Larkin JC, et al. Perinatal Outcomes and Unconventional Natural Gas Operations in Southwest Pennsylvania. PLOS ONE 2015;10(6):e0126425. DOI: 10.1371/journal.pone.0126425.
6. Casey JA, Savitz DA, Rasmussen SG, et al. Unconventional natural gas development and birth outcomes in Pennsylvania, USA. Epidemiology 2016;27(2):163-72. (In eng). DOI: 10.1097/ede.0000000000000387.
7. Ma Z-Q. Time Series Evaluation of Birth Defects in Areas with and without Unconventional Natural Gas Development. Journal of Epidemiology and Public Health Reviews 2016;1. DOI: 10.16966/2471-8211.107.
8. Rasmussen SG, Ogburn EL, McCormack M, et al. Association Between Unconventional Natural Gas Development in the Marcellus Shale and Asthma Exacerbations. JAMA Internal Medicine 2016;176(9):1334-1343. DOI: 10.1001/jamainternmed.2016.2436.
9. Stacy S. A Review of the Human Health Impacts of Unconventional Natural Gas Development. Current Epidemiol Review 2017;4(1). DOI: doi.org/10.1007/s40471-017-0097-9.
10. Whitworth KW, Marshall AK, Symanski E. Maternal residential proximity to unconventional gas development and perinatal outcomes among a diverse urban population in Texas. PLoS One 2017;12(7):e0180966. (In eng). DOI: 10.1371/journal.pone.0180966.
11. Tustin Aaron W, Hirsch Annemarie G, Rasmussen Sara G, Casey Joan A, Bandeen-Roche K, Schwartz Brian S. Associations between Unconventional Natural Gas Development and Nasal and Sinus, Migraine Headache, and Fatigue Symptoms in Pennsylvania. Environmental Health Perspectives 2017;125(2):189-197. DOI: 10.1289/EHP281.
12. Walker Whitworth K, Kaye Marshall A, Symanski E. Drilling and Production Activity Related to Unconventional Gas Development and Severity of Preterm Birth. Environ Health Perspect 2018;126(3):037006. (In eng). DOI: 10.1289/ehp2622.
13. Koehler K, Ellis JH, Casey JA, et al. Exposure Assessment Using Secondary Data Sources in Unconventional Natural Gas Development and Health Studies. Environmental Science & Technology 2018;52(10):6061-6069. DOI: 10.1021/acs.est.8b00507.

14. Lewis C, Greiner LH, Brown DR. Setback distances for unconventional oil and gas development: Delphi study results. PLOS ONE 2018;13(8):e0202462. DOI: 10.1371/journal.pone.0202462.
15. Casey JA, Goin DE, Rudolph KE, et al. Unconventional natural gas development and adverse birth outcomes in Pennsylvania: The potential mediating role of antenatal anxiety and depression. Environ Res 2019;177:108598. (In eng). DOI: 10.1016/j.envres.2019.108598.
16. Deziel NC, Brokovich E, Grotto I, et al. Unconventional oil and gas development and health outcomes: A scoping review of the epidemiological research. Environmental Research 2020;182:109124. DOI: https://doi.org/10.1016/j.envres.2020.109124.
17. Gonzalez DJX, Sherris AR, Yang W, et al. Oil and gas production and spontaneous preterm birth in the San Joaquin Valley, CA: A case-control study. Environ Epidemiol 2020;4(4):e099. (In eng). DOI: 10.1097/EE9.0000000000000099.
18. Willis MD, Hill EL, Boslett A, Kile ML, Carozza SE, Hystad P. Associations between Residential Proximity to Oil and Gas Drilling and Term Birth Weight and Small-for-Gestational-Age Infants in Texas: A Difference-in-Differences Analysis. Environ Health Perspect 2021;129(7):77002. (In eng). DOI: 10.1289/EHP7678.
19. Deziel NC, Clark CJ, Casey JA, Bell ML, Plata DL, Saiers JE. Assessing Exposure to Unconventional Oil and Gas Development: Strengths, Challenges, and Implications for Epidemiologic Research. Current Environmental Health Reports 2022;9(3):436-450. DOI: 10.1007/s40572-022-00358-4.
20. Finkel ML, Hays J. Environmental and health impacts of 'fracking': why epidemiological studies are necessary. J Epidemiol Community Health 2016;70(3):221-2. (In eng). DOI: 10.1136/jech-2015-205487.
21. Wilke RA, Freeman JW. Potential Health Implications Related to Fracking. JAMA 2017;318(17):1645-1646. (In eng). DOI: 10.1001/jama.2017.14239.
22. Currie J, Greenstone M, Meckel K. Hydraulic fracturing and infant health: New evidence from Pennsylvania. Science Advances 2017;3(12):e1603021. DOI: 10.1126/sciadv.1603021.
23. McKenzie LM, Guo R, Witter RZ, Savitz DA, Newman LS, Adgate JL. Birth Outcomes and Maternal Residential Proximity to Natural Gas Development in Rural Colorado. Environ Health Perspect2014:412-7.
24. McKenzie LM, Allshouse W, Daniels S. Congenital heart defects and intensity of oil and gas well site activities in early pregnancy. Environment International 2019;132:104949. DOI: https://doi.org/10.1016/j.envint.2019.104949.
25. The Process of Unconventional Natural Gas Production. EPA.
26. Webb E, Hays J, Dyrszka L, et al. Potential hazards of air pollutant emissions from unconventional oil and natural gas operations on the respiratory health of children and infants. Rev Environ Health 2016;31(2):225-43. (In eng). DOI: 10.1515/reveh-2014-0070.
27. Willis MD, Jusko TA, Halterman JS, Hill EL. Unconventional natural gas development and pediatric asthma hospitalizations in Pennsylvania. Environ Res 2018;166:402-408. (In eng). DOI: 10.1016/j.envres.2018.06.022.
28. Holguin F, Cardet JC, Chung KF, et al. Management of Severe Asthma: a European Respiratory Society/American Thoracic Society Guideline. European Respiratory Journal 2019:1900588. DOI: 10.1183/13993003.00588-2019.

29. Nuttall FQ. Body Mass Index: Obesity, BMI, and Health: A Critical Review. Nutrition today 2015;50(3):117-128. (In eng). DOI: 10.1097/NT.0000000000000092.