Environmental Science & Technology

Article

Cite This: Environ. Sci. Technol. 2018, 52, 6061−6069

pubs.acs.org/est

Exposure Assessment Using Secondary Data Sources in Unconventional Natural Gas Development and Health Studies

Kirsten Koehler,[†] J. Hugh Ellis,[†] Joan A. Casey,[‡] David Manthos,[§] Karen Bandeen-Roche,[∥] Rutherford Platt,[⊥] and Brian S. Schwartz*[,†,#,∇]

[†]Department of Environmental Health and Engineering, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland 21205, United States

[‡]Division of Environmental Health Sciences, University of California at Berkeley School of Public Health, Berkeley, California 94720, United States

[§]SkyTruth, Shepherdstown, West Virginia 25443, United States

[∥]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland 21205, United States

[⊥]Department of Environmental Studies, Gettysburg College, Gettysburg, Pennsylvania 17325, United States

[#]Department of Epidemiology and Health Services Research, Geisinger Health System, Danville, Pennsylvania 17822, United States

[∇]Department of Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, United States

Supporting Information

ABSTRACT: Studies of unconventional natural gas development (UNGD) and health have ranked participants along a gradient of geographic information system (GIS)-based activity that incorporated the distance between participants' home addresses and unconventional natural gas wells. However, studies have used different activity metrics, making result comparisons across the studies difficult. The existing studies have only incorporated wells, without accounting for other components of development (e.g., compressors, impoundments, and flaring events), for which it is often difficult to obtain reliable data but may have relevance to health. Our aims were to (1) describe, in space and time, UNGD-related compressors, impoundments, and flaring events; (2) evaluate whether and how to incorporate these into UNGD activity assessment; and (3) evaluate associations of these different approaches with mild asthma exacerbations. We identified 361 compressor stations, 1218 impoundments, and 216 locations with flaring events. A principal component analysis identified a single component that was approximately an equal mix of the metrics for compressors, impoundments, and four phases of well development (pad preparation, drilling, stimulation, and production). However, temporal coverage for impoundments and flaring data was sparse. Ultimately, we evaluated three UNGD activity metrics, including two based on the existing studies and a novel metric that included well pad development, drilling, stimulation, production, and compressor engine aspects of UNGD. The three metrics had varying magnitudes of association with mild asthma exacerbations, although the highest category of each metric (vs the lowest) was associated with the outcome.

INTRODUCTION

Unconventional natural gas (UNG) constitutes over 40% of the natural gas produced in the U.S., up from less than 10% in 2007. Pennsylvania's Marcellus shale accounts for over a quarter of the country's UNG production.[1] Several epidemiology studies evaluated associations of unconventional natural gas development (UNGD) with health outcomes, but these studies used different UNGD metrics to categorize participants, making comparing results difficult, and these metrics have only incorporated wells, though wells are just one component of UNGD-related infrastructure.

New UNGD involves pad preparation, drilling, perforation, stimulation, and gas production. The fluid returning to the surface with the gas can be stored in surface impoundments, where volatile organic compounds (VOCs) evaporate. Gas is compressed using diesel or natural gas powered compressor engines before distribution.[2] Residents of regions undergoing UNGD face potential chemical exposures, through water, soil, and air; physical agent exposures, including noise, light, and vibration; and community impacts.[2−21] Exposure to UNGD is not a single exposure, but multiple time varying exposures, each with different scales of impact.

Epidemiologic studies have evaluated the associations of UNGD with birth outcomes,[22−25] asthma symptoms,[26−28] cancer,[29,30] hospitalization rates,[31] and car crashes.[5] In our prior

Received: January 26, 2018
Revised: April 20, 2018
Accepted: April 26, 2018
Published: April 26, 2018

DOI: 10.1021/acs.est.8b00507
Environ. Sci. Technol. 2018, 52, 6061−6069

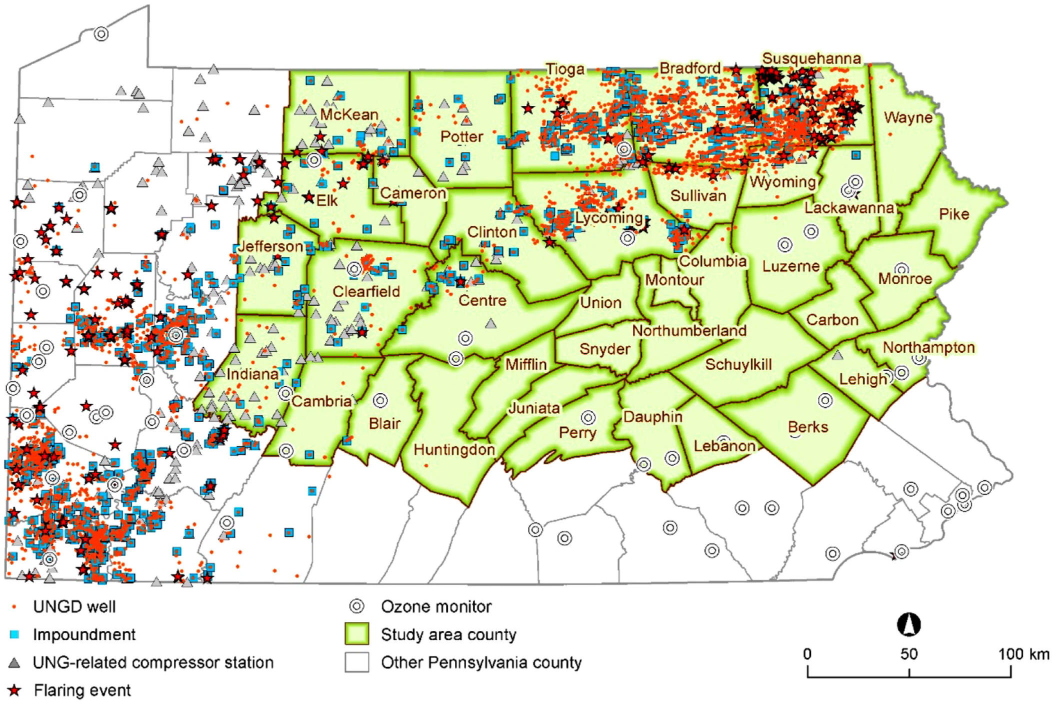

Figure 1. Location of UNG-related impoundments, compressor engines, and UNG wells. Impoundments included those identified in 2005, 2008, 2010, and 2013 (n = 1,218); compressor engines included those started by 2013 (n = 861); and wells included those drilled by 2015 (n = 9669). The counties in green are those that were included in the fishnet grid. Ozone monitors (n = 55) are those that were active in 2012. UNGD, Unconventional natural gas development and UNG, unconventional natural gas.

study, we evaluated the associations of four phases of UNG well development with mild, moderate, and severe asthma exacerbations, among 35 508 primary care patients with asthma of the Geisinger Clinic in Pennsylvania from 2005 to 2012.[32] The epidemiologic studies that assigned UNGD metrics on an individual level[22−25,27,28,32] assigned their metrics using a geographic information system (GIS)-based proxy that incorporated the distance between study participants' home addresses and UNG wells. Studies have used different approaches to defining GIS-based metrics, limiting comparability. In addition, GIS-based proxies are crude measures of UNGD-related exposures, but they work well retrospectively and are inexpensive compared to multiple pathway exposure assessment of physical, chemical, and social impacts.

To date, epidemiologic studies have only incorporated wells into exposure metrics, even though other components of UNGD, such as compressor stations and impoundments, may be major contributors to air emissions.[13,21] Because no prior study has attempted to incorporate impoundments and compressor engines into UNGD metrics, it is not clear what (if anything) they add to metric creation. Additionally, no study has incorporated flaring events, which are sources of combustion products and VOCs.

To address the limitations related to UNGD metrics used in epidemiology studies, the three primary aims of the analyses in this article were to (1) characterize UNGD-related impoundments, compressor engines, and flaring events in Pennsylvania; (2) evaluate whether and how to incorporate impoundments, compressor engines, and flaring events into UNGD activity metrics; and (3) compare associations of different GIS-based UNGD metrics used in existing studies to each other and in their associations with mild asthma exacerbations.

METHODS

UNGD-Related Compressor Engines, Impoundments, and Flaring Events in Pennsylvania. Compressor stations pressurize the gas to keep it flowing through pipelines. The number of engines required at a compressor station and the operating conditions of the engines depends on the pressure of the gas received by the station. Unlike data on wells,[33,34] data on compressor engines and impoundments are not available electronically. To identify compressor engines, we obtained a list of compressor stations thought to be UNGD-related from the Pennsylvania Department of Environmental Protection (DEP) (n = 506). We visited four DEP locations (Northeast, North-central, Northwest, and Southwest) and scanned relevant documents (including applications, general information forms, authorizations, start letters, and cancellations; n = 6,007) between October 2013 and May 2014. We data abstracted these documents for station name, location, number of compressor engines, compressor engine horsepower, compressor engine emissions, expected start date of operation, authorization date, start date, and cancellation date; 2700 documents contained at least one of these variables (initially, we scanned unnecessary documents; later, we refined the process on which documents to scan). We excluded compressor stations that had no available documents or, upon document review, were not UNGD-related (n = 49). After data entry, we did data checking to confirm the accuracy of the entered data. We used compressor station names and site identification numbers to link data across the documents. If, say, horsepower was missing in one document, we looked for it in other documents for that compressor engine.

Information on impoundment location and sizes was obtained in partnership with SkyTruth, which created a collaborative image analysis application on their Web site that

DOI: 10.1021/acs.est.8b00507
Environ. Sci. Technol. 2018, 52, 6061−6069

displayed aerial imagery collected by the USDA National Agricultural Imagery Program[35] of the one square kilometer area around UNG wells from the summers of 2005, 2008, 2010, and 2013 (Abstract graphic). We used the Breaks for Additive Season and Trend package in *R* to identify the direction, magnitude, and timing of time series breaks in impoundments (details are provided in the Supporting Information).[36,37] Based on these breaks we identified the approximate dates of creation and removal of the impoundments.

We also identified flaring events using detections recorded at night by the Visible Infrared Imaging Radiometer Suite on the Suomi NPP satellite operated by the National Oceanic and Atmospheric Administration (NOAA). Methods related to the identification of flaring events are available in the Supporting Information.

Incorporate Impoundments and Compressor Engines into UNGD Activity Assessment. Principle components analysis (PCA) is a mathematical transformation process that retains important trends and patterns in data, while reducing the number of dimensions. The goal of PCA is to find components, linear combinations of the different variables in the data set, that explain the highest proportion of the variability in the data set.[38] We used PCA to assess the relationship between metrics created for four phases of well development (pad preparation, drilling, stimulation, and production[22,27,32]), compressor engines, and impoundments. We did not incorporate flaring events into the PCA because we did not have information on flaring events before 2013, and only four locations had flaring events identified in 2013. We created a regular grid (5 × 5 km) across 38 counties in central and northeastern Pennsylvania (Figure 1, in green) (number of grid points = 2627). The 5 × 5 grid was a compromise between spatial resolution and computational requirements, in addition to being a reasonable scale for regional air pollutants. We used this regular grid to assign exposure metrics, instead of using residential locations of Geisinger patients, so that the locations of the points included in the PCA would not be affected by residential patterns or population density. Although there was still a spatial correlation structure between the grid points included in the PCA, we aimed primarily to build an index rather than to study correlation structure, and thus we did not consider this a major limitation. On January 1 and July 1 for 2005–2013 (18 time points), we assigned inverse distance-squared (IDW_2) development and infrastructure metrics to each grid point for four phases of well development, impoundments, and compressor engines as follows (eq 1):

$$\text{Development or infrastructure metric for grid point j} = \sum_{i=1}^{m} \frac{s_i}{d_{ij}^2} \tag{1}$$

For each IDW_2 metric, m was either the number of wells in the given phase, started compressor engines, or installed impoundments; and d_{ij}^2 was the squared-distance (meters) between the well, compressor engine, or impoundment i and grid centroid j. For the four phases of well development, s_i was 1 for the pad production and drilling phases, total well depth (meters) of well i for the stimulation phase, and daily natural gas production volume (m^3) of well i for the production phase. For compressor engines, s_i was the compressor engine horsepower. Engines contributed to the metric from their start date to their removal date. For impoundments, s_i was the area (m^2) of the impoundment, which contributed to the metric from their installation to their removal date. For years with aerial imagery (2005, 2008, 2010, and 2013), we assigned six development and infrastructure metrics (impoundments, compressor engines, and four phases of well development). For the remaining dates, we assigned all development and infrastructure metrics, except the one for impoundments (i.e., compressor engines and four phases of well development). On some dates, there were no wells in a given phase, so that phase's metric was not included in the PCA for that date.

On each of the 18 dates evaluated, we truncated the UNGD metrics at their 98th percentile, log and z transformed the truncated values to normalize distributions, put the development and infrastructure metrics on the same scale, and conducted a PCA using the Pearson correlation matrix in Stata. We compared loadings and scree plots across the evaluated dates. We also compared the first component from the PCA to a summed *z*-score of all UNGD development and infrastructure metrics available on that date.

Comparison of GIS-Based Metrics and Their Associations with Mild Asthma Exacerbations. We compared how the different approaches to UNGD metrics categorized patient assignments of the UNGD activity by evaluating the sensitivity of their associations with mild asthma exacerbations (identified by new oral corticosteroid [OCS] medication orders for asthma) compared to the control encounters. Data on asthma exacerbations and control patient contact dates with the health system came from the Geisinger electronic health record.[32] Geisinger provides primary care to over 400 000 patients in Pennsylvania and New York, and the primary care population is a representative sample of the general population. New OCS orders for an asthma exacerbation were identified, and the medication order date was considered the index date. Controls were identified from the population of patients with asthma under observation by Geisinger and frequency matched to case patients by age category (5–12, 13–18, 19–44, 45–61, 62–74, or ≥75 years), sex (male or female), and year of encounter.

To compare how different the UNGD activity metrics categorized subjects on UNGD, here, we assigned three UNGD metrics (described below) to asthma case and control encounters identified in a previous study.[32] We then compared how each metric ranked the case and control encounter dates using Spearman correlations for continuous metrics and tables for categorical metrics. Next, we evaluated associations of each of these metrics with mild asthma exacerbations (n = 39 442 case and control dates for 13196 cases and 18693 controls) using our previously reported adjusted multilevel model.[32]

We evaluated three different approaches to UNGD metrics: (1) the categorical distance to the nearest drilled well (DNDW); (2) the inverse distance metric based on the drilling phase (IDW_1); and (3) the inverse distance-squared metric (IDW_2) incorporating the four phases of well development and compressor engines ($IDW_2$4C). The DNDW approach, used by Rabinowitz,[28] was based on the distance from a patient's home to the closest drilled well of any age, and it was categorized into less than 1 km, 1–2 km, and greater than 2 km. The IDW_1 metric, similar to that used by McKenzie and Stacy,[24,25] was assigned as follows (eq 2):

$$IDW_1 \text{ metric} = \sum_{i=1}^{n} \frac{1}{d_{ij}} \tag{2}$$

DOI: 10.1021/acs.est.8b00507
Environ. Sci. Technol. 2018, 52, 6061–6069

Table 1. Results of PCA with Percentage of Variation Explained by Component 1 and Component 1 Loadings

date	proportion of variance explained by component 1	compressor engine metric	component 1 loadings: well metrics: pad	drilling	stimulation	production	impoundment metric	correlation of component 1 with z-score
1/1/2005	0.77	0.50	[a]	[a]	[a]	0.62	0.59	0.99
7/1/2005	0.76	0.47	0.46	0.33	[a]	0.47	0.49	0.99
1/1/2006	0.91	0.56	0.59	[a]	[a]	0.58	[b]	0.99
7/1/2006	0.94	0.42	0.46	0.46	0.46	0.44	[b]	0.99
1/1/2007	0.85	0.46	0.37	0.45	0.47	0.47	[b]	0.99
7/1/2007	0.72	0.50	0.21	0.50	0.45	0.51	[b]	0.99
1/1/2008	0.72	0.46	0.36	0.43	0.43	0.46	0.30	0.99
7/1/2008	0.58	0.46	0.35	0.34	0.43	0.48	0.36	0.99
1/1/2009	0.58	0.47	0.53	0.41	0.47	0.32	[b]	0.99
7/1/2009	0.69	0.24	0.50	0.50	0.48	0.46	[b]	0.99
1/1/2010	0.67	0.33	0.36	0.45	0.39	0.45	0.45	0.99
7/1/2010	0.81	0.34	0.43	0.43	0.42	0.40	0.42	0.99
1/1/2011	0.80	0.38	0.47	0.46	0.46	0.46	[b]	0.99
7/1/2011	0.83	0.41	0.46	0.46	0.45	0.46	[b]	0.99
1/1/2012	0.84	0.41	0.44	0.46	0.46	0.46	[b]	0.99
7/1/2012	0.79	0.40	0.43	0.47	0.45	0.48	[b]	0.99
1/1/2013	0.83	0.38	0.41	0.42	0.42	0.43	0.39	0.99
7/1/2013	0.78	0.41	0.37	0.41	0.41	0.44	0.40	0.99

[a]All grid points had a value of zero for this variable on this date, and variables with zero variance were dropped from PCA. [b]Impoundment data was only available in 2005, 2008, 2010, and 2013.

In eq 2, n was the number of drilled wells within 10 miles of a patient's home, j, and d_{ij} was the distance between a well and the patient's home. We tertiled the IDW_1 metric using case and control encounters with at least one well within 10 miles and created a reference group of case and control encounters with no wells within 10 miles, as McKenzie et al. did in their Colorado birth outcome study.[24] The $IDW_2$4C metric included the development and infrastructure metrics from the four phases of well development and UNG-related compressor engines. As described above (eq 1), we assigned each encounter date a value for four phases of well development and compressor stations, created z-scores for each of the five values, summed the z-scores, and quartiled the sum using all of the patient events (exacerbations or control dates). The results of the PCA (see the Results section) informed the creation of the $IDW_2$4C metric. We did not include impoundments in the final PCA model used to assign the UNGD activity to subjects because impoundment data were not available for all of the years.

To evaluate the sensitivity of associations of different approaches to UNGD metric creation with a health outcome, we evaluated the associations of the DNDW, IDW_1, and $IDW_2$4C metrics with mild asthma exacerbations using multilevel logistic regression with a random intercept for patient and community to account for multiple events per patient and clustering in communities. These models were adjusted for potential confounding variables as previously described,[32] including age, sex, race/ethnicity, family history of asthma, smoking status, season, medical assistance, overweight/obesity status, distance and distance-squared to nearest major and minor arterial roads, maximum temperature and maximum temperature-squared on the day prior to the event, and community socioeconomic deprivation.[39−41] We then compared the odd ratios from each of these models. The study was approved by the Geisinger Institutional Review Board (with an IRB authorization agreement with Johns Hopkins Bloomberg School of Public Health).

RESULTS

UNGD-Related Compressor Engines, Impoundments, and Flaring Events in Pennsylvania. We identified 1218 impoundments and 457 compressor stations in Pennsylvania (Figures 1 and S1). The median areas (m^2) of impoundments in 2005, 2008, 2010, and 2013 were 344.0, 558.8, 1990.2, and 6209.7, respectively. The average estimated duration of an impoundment from installation to removal was 1.9 years. At the 457 compressor stations, we identified 1419 compressor engines (maximum of 20 engines at a single station), though only 861 engines at 361 stations had start letters stating they were operational. The date of development for compressor engines and impoundments was similar to that for wells. Although the number of impoundments decreased from 2010 to 2013, the total area of impoundments increased from 1.96 to 3.96 km^2. We identified flares at 216 locations (Figure 1) between September 2012 and August 2015, but these data were not used in further analyses because they were not available for a majority of the study period. Access to the impoundment and flaring data is described in the Supporting Information.

PCA Applied to Wells, Compressor Stations, and Impoundments. PCA reduces the dimensionality of data (i.e., the number of variables) by identifying a small number of components. These components extract the maximum variance in a data set by analyzing the total variance.[42] In each PCA, the first component explained between 58 and 94% (median 79%) of the total variation (Table 1). For 15 of the 18 dates, only the first component had an eigenvalue above one.[43] The first components' loadings were consistently made up of an approximately equal mix of the UNGD metrics, suggesting that the four phases of well development, compressor stations, and impoundments generally co-occur in space and time. Each of these measures was an important measure of UNGD activity.

DOI: 10.1021/acs.est.8b00507
Environ. Sci. Technol. 2018, 52, 6061−6069

We left impoundments out of the final combined metric, however, because information on impoundment ponds were not available for all of the years for which we had asthma data. The first component was also highly correlated with a summed z-score of the metrics on each date (Spearman correlations >0.99). In contrast, the second component, which explained between 4 and 29% of the variation, did not have consistent loadings, although the compressor metric tended to be the largest (Table S1). Based on these results and because the single component ranked points similarly to a z-score of the four well phases plus compressor station metrics, we created the $IDW_2$4C metric by summing the z-score of these five metrics.

Comparison of GIS-based UNGD activity metrics. We sought to compare how the DNDW, IDW_1, and $IDW_2$4C metrics ranked the participant index dates along a UNGD activity gradient. We then categorized these continuous UNGD activity metrics into tertiles or quartiles, consistent with prior research use.[22,24,25,27,28] Comparing the DNDW and $IDW_2$4C metrics (Table 2), 96.4% of the participant index dates in the $IDW_2$4C metric's highest quartile were also in the highest category of the DNDW metric (greater than 2 km from the closest well), but 98.6% of the index dates in the $IDW_2$4C metric's highest category were greater than 2 km from the closest well. For the IDW_1 and $IDW_2$4C metrics, we compared both the continuous and categorical metrics. The Spearman correlation for continuous IDW_1 and $IDW_2$4C metrics was 0.36. While 80.3% of the assignments for the IDW_1 metric's highest tertile were also in the highest quartile of $IDW_2$4C, 18.5% of the assignments for IDW_1's lowest category (no wells within 10 miles) were in $IDW_2$4C's highest quartile (Table 3).

Table 2. Categorization of Case and Control Encounter Dates (counts) by Distance to Nearest Drilled Well (DNDW) and by an Inverse Distance Squared Metric Incorporating Four Phases of Well Development and Compressor Engines ($IDW_2$4C)

		DNDW categories[a]			
		<1 km	1−2 km	>2 km	total
$IDW_2$4C categories[b]	Q1[c]	2	4	17 381	17 387
	Q2	4	30	17 353	17 387
	Q3	4	46	17 337	17 387
	Q4	238	385	16 764	17 387
	total	248	465	68 835	69 548

[a]Distance to the nearest drilled well, based on Rabinowitz.[28] [b]An inverse distance metric incorporating four phases of well development (pad preparation, drilling, stimulation, and production) and UNG-related compressor stations, based on Casey, Tustin, and Rasmussen.[32] [c]Quartile.

We then compared associations of the DNDW, IDW_1, and $IDW_2$4C metrics with a health outcome, mild asthma exacerbations. In these models, the highest group of each metric (vs the lowest) was associated with increased odds of mild exacerbation, though the magnitudes of association differed with $IDW_2$4C the most (IDW_1< DNDW < $IDW_2$4C, Table 4). The DNDW and $IDW_2$4C metrics had increasing odds ratios across UNGD categories, whereas the second tertile for the IDW_1 metric had a slightly stronger association with the outcome than that for the third tertile. Associations of the $IDW_2$4C metric with mild asthma exacerbations were intermediate of those from four regressions of each phase of well development separately in our prior study.[32] In this example, we observed positive associations between each UNGD metric and mild asthma exacerbations. The highest level of exposure for the newly developed $IDW_2$4C metric, however, exhibited a magnitude of association twice that of the previously reported IDW_1 and DNDW metrics. In addition, both the DNDW and the $IDW_2$4C metrics exhibited exposure-response trends across quantiles of exposure, while the IDW_1 metric did not. Finally, we compare the results to our previous work for the production volume only (IDW_2P).[32] The OR for mild exacerbations in the highest quartile of IDW_2P exposure was larger than that for the $IDW_2$4C metric.

Table 3. Categorization of Case and Control Encounter Dates (Counts) by an Inverse Distance Metric That Was Based only on the Drilling Phase Inverse Distance (IDW_1) and an Inverse Distance Squared Metric Incorporating Four Phases of Well Development and Compressor Engines ($IDW_2$4C)

		IDW_1 tertiles[a]				
		0 wells in 10 miles	T1[b]	T2	T3	total
$IDW_2$4C quartiles[c]	Q1[d]	16 999	159	146	83	17 387
	Q2	15 158	954	965	310	17 387
	Q3	14 866	1050	1086	385	17 387
	Q4	10 649	1796	1762	3180	17 387
	total	57 672	3959	3959	3959	69 548

[a]An inverse distance metric incorporating drilled unconventional wells. [b]Tertile. [c]An inverse distance metric incorporating four phases of well development (pad preparation, drilling, stimulation, and production) and UNG-related compressor stations, based on Casey, Tustin, and Rasmussen.[32] [d]Quartile.

DISCUSSION

Compressor engines, impoundments, and flaring events are potential sources of emissions related to UNGD that have not previously been described or incorporated in epidemiology studies, in part because data are not readily available. The value of including these additional sources of information on UNGD, particularly in health studies, remains unknown. Additionally, approaches to incorporating wells into exposure metrics have differed across epidemiology studies. In the presented work, we aimed to systematically collect data on a range of different components of UNGD activity, compared several ways of computing activity metrics, and evaluated associations of these metrics with an important and prevalent health outcome. We first identified locations and dates of UNGD-related compressor engines, impoundments, and flaring events in Pennsylvania. A PCA reduced the number of UNGD activity variables and summarized meaningful groupings of compressor stations, impoundments, and the four phases of UNG well development for activity assessment. Based on the PCA, we created a new activity metric (i.e., $IDW_2$4C), a summed z-score of the four well plus compressor station metrics. $IDW_2$4C exhibited stronger associations with mild asthma exacerbations than did the DNDW[28] and IDW_1[24,25] metrics.

Despite significant inputs of time, the utility of adding the information on flaring and impoundments for UNGD activity assessment was unclear. Flaring information was so sparse that it could not be included in the PCA. The PCA suggested that the four well phases, compressors, and impoundments co-occur in space and time, and that the underlying latent construct of

DOI: 10.1021/acs.est.8b00507
Environ. Sci. Technol. 2018, 52, 6061−6069

Table 4. Associations of Unconventional Natural Gas Development (UNGD) Metrics and with Mild Asthma Exacerbations[a]

UNGD metric included in model[b]	category	odds ratio (95% CI[c])
DNDW[d]	>2 km (REF)	1.0
	1−2 km	1.13 (0.76−1.69)
	<1 km	1.83 (1.03−3.25)
IDW_1[e]	no wells within 10 miles (REF)	1.0
	tertile 1	0.96 (0.83−1.13)
	tertile 2	1.21 (1.03−1.42)
	tertile 3	1.19 (1.01−1.41)
$IDW_2$4C[f]	quartile 1 (REF)	1.0
	quartile 2	1.31 (1.16−1.48)
	quartile 3	2.20 (1.93−2.52)
	quartile 4	3.69 (3.16−4.30)
Results from Prior Published Analysis in Ref 32		
IDW_2P[g]	quartile 1 (REF)	1.0
	quartile 2	1.28 (1.13−1.46)
	quartile 3	2.15 (1.87−2.47)
	quartile 4	4.43 (3.75−5.22)

[a]New oral corticosteroid medication orders. [b]Multilevel models with a random intercept for patient and community, adjusted for age category (5−12, 13−18, 19−44, 45−61, 62−74, and 75+ years), sex (male, female), race/ethnicity (white, black, Hispanic, and other), family history of asthma (yes vs no), smoking status (never, former, current, and missing), season (spring, March 22−June 21; summer, June 22−September 21; fall, September 22−December 21;and winter, December 22−March 21), medical assistance (yes vs no), overweight/obesity (normal, body mass index [BMI] < 85th percentile or BMI < 25 kg/m^2; overweight, BMI = 85th to <95th percentile or BMI = 25 to <30 kg/m^2; obese, BMI ≥ 95th percentile or BMI ≥ 30 kg/m^2, for children and adults, respectively; and BMI missing), type 2 diabetes (yes vs no), community socioeconomic deprivation (quartiles), distance to nearest major and minor arterial road (truncated at the 98th percentile, meters, and z-transformed), squared distance to the nearest major and minor arterial road (truncated at the 98th percentile, meters, and z-transformed), maximum temperature on the day prior to the event (degrees Celsius), and squared maximum temperature on the day prior to the event (degrees Celsius). [c]Confidence interval. [d]Distance to the nearest drilled well, based on Rabinowitz.[28] [e]An inverse distance metric that was based only on the drilling phase. [f]An inverse distance-squared metric incorporating four phases of well development (pad preparation, drilling, stimulation, and production) and UNG-related compressor engines, based on Casey, Tustin, and Rasmussen.[32] [g]An inverse distance squared metric that was based only on production volume. See Rasmussen et al.[32] for full details.

"UNGD activity" can be adequately measured without investment of time and resources to measure all six. However, the major limitation of compressor and impoundment information was access and availability, and if this information become more systematically and widely available, it could help improve UNGD activity assessment. In addition to a high proportion of missing data, flaring and impoundment data were likely measured with more error than were data for wells. We likely underestimated counts of impoundments because we only had aerial imagery for four years between 2005 and 2013. Because the average estimated duration of an impoundment from installation to removal was 1.9 years, there were likely impoundments that were installed and removed in between the years with images, and thus they would not have been in our data set. Additionally, we did not look for impoundments that were more than 1 km from a well. Data on flares were not available throughout the study period, ability to identify flares was hampered by cloud cover, and we may have missed shorter-duration flares due to frequency of Suomi NPP satellite passes. While impoundments were included in the PCA, both impoundments and flares were not included in the new $IDW_2$4C metric because of missing data.

The PCA suggested that on a majority of days evaluated, a single component captured most, but not all, of the variation of the compressor engine, impoundment, and well IDW_2 metrics. It was not unexpected that the PCA loaded on a single component, since the wells, impoundments, and compressor engines had similar spatial and temporal distributions. We opted to create a z-score of the four well phases plus compressor station metrics to improve reproducibility in future studies and because of the strong correlation between the first PCA component and z-score.

We observed substantial differences in the DNDW, IDW_1, and $IDW_2$4C metrics, which were originally designed for studies conducted in different regions and time periods, and how they ranked in participant case and control index dates. For example, only 18.3% of participant index dates in the $IDW_2$4C metric's highest quartile were in the highest tertile of the IDW_1 metric. The DNDW metric was used in a study in southwestern Pennsylvania in 2012,[28] and the IDW_1 metric was used in a study in Colorado from 1996 and 2009.[24] The participants included in those studies lived, on average, closer to wells than in the studies that used the IDW_2 metric, which were conducted from 2005 to 2015 in central and northeastern Pennsylvania.[22,27,32] This was, in part, because the southwestern Pennsylvania study did not include the earlier years of UNGD, when wells were less dense, and the Colorado study included both conventional and UNG wells.

Each of the three metrics also incorporated different information. Because the DNDW metric was categorized based on distance to the single closest drilled well, it did not take into account the size, number, or density of wells. The DNDW and IDW_1 metrics incorporated all existing wells that had been directionally drilled and/or hydraulically fractured,[24,25,28] whereas the $IDW_2$4C distinguished between four phases of development. Both the DNDW and IDW_1 metrics assumed that all exposures from wells were continuous after a well was drilled, and that exposures were equal from all drilled wells, regardless of phase of development, depth of the well, or volume of natural gas produced at the well. However, phase of development is important to incorporate into metric formulation because exposures, such as air emissions, differ by phase of development, and because not all drilled wells are later stimulated or produce natural gas. Of the 9669 unconventional natural gas wells drilled in Pennsylvania by 2015, 1992 did not have stimulation dates, and of wells with stimulation dates, 377 did not report production (although it is not possible to distinguish between missing and zero values in the data). The DNDW metric assumed that wells farther than 2 km, and the IDW_1 assumed wells farther than 10 miles, from a participant's home did not contribute to exposure from UNGD activity, assumptions that may not be true for exposures such as regional air pollutants (e.g., ozone and particulate matter). No UNGD metric took into consideration the full variability of potential exposures. For example, safety practices (and potential accidental exposures) differ between well operators,[44] and impacts also vary over time, as regulations are enacted and industry practices change.

DOI: 10.1021/acs.est.8b00507
Environ. Sci. Technol. 2018, 52, 6061−6069

We designed our $IDW_2 4C$ metric to capture many potential exposure pathways associated with UNGD. The $IDW_2 4C$ metric assumed that wells only contributed to UNGD activity during the four phases of development (pad preparation, drilling, stimulation, or production), and in between these phases, wells did not contribute to the metric. The $IDW_2 4C$ metric also assumed wells contributed differently to activity during the stimulation phases (proportional to total depth) and production (proportional to gas quantity produced) phases, and compressor engines contributed differently (proportional to total horsepower). We hypothesized that these were reasonable assumptions because well depth should be a surrogate, for example, for the number of truck trips to well pads bringing pipe, chemicals, and water used in stimulation,[45,46] daily volume of natural gas production should be a surrogate for activity and possibly for fugitive emissions. However, we acknowledge that without environmental measurements, we cannot definitively say how well our metric captures each potential exposure pathway. In addition, our metric, like others, only accounted for UNGD activity assignments at a 5 × 5 km grid cell containing the participants' homes, which may have resulted in bias.[47] We were unable to estimate contact with UNGD at other locations that may result from occupation, transportation, exercise, or other activities of daily living.[48] Rigorous approaches to characterize each of the potential chemical, physical, and social exposures from UNGD exist,[17,49−51] but such approaches may not work well retrospectively and are much more time-consuming and costly than GIS-based approaches.

We compared the associations of the different UNGD metrics with mild asthma exacerbations. Although inference was the same across the three metrics (the highest group of each was associated with mild asthma exacerbations), the magnitude of the odds ratios and the trends across activity categories differed substantially. The $IDW_2 4C$ metric was most strongly associated with mild asthma exacerbations, and the IDW_1 metric was the least strongly associated. However, the association was not as strong for the $IDW_2 4C$, as we previously found for a metric that only incorporated the production phase. Each metric represents different data requirements. The IDW_1 and DNDW metrics have low data requirements, needing only the locations of wells. The $IDW_2 P$ metric has somewhat increased data needs, additionally requiring the production volume over time. Finally, the $IDW_2 4C$ metric has substantial data requirements, much of which is not easily accessible digitally. Had we used the IDW_1 or DNDW metric in our original study, we would have come to different conclusions on the strength of the association of UNGD and asthma exacerbations. Because the associations of the $IDW_2 4C$ metric with mild asthma exacerbations were intermediate of those from each phase of well development separately (as in our prior study),[32] we concluded that the time, effort, and expense required to collect and data enter information on compressor engines did not substantively change interpretation of the associations with mild asthma exacerbations reported herein and in the prior paper.[32] It is possible that the DNDW and IDW_1 metrics had more misclassification than the $IDW_2 4C$ metric, which could explain the larger magnitude association between $IDW_2 4C$ and mild asthma exacerbations. Without environmental measurements, however, we cannot quantify how well each metric captures potential exposures from UNGD, and so we cannot definitively interpret the decrease in magnitude of the association with the DNDW and IDW_1 metrics compared to the $IDW_2 4C$ metric. We also acknowledge that we cannot rule out the potential for unmeasured confounding in each model, just as we could not do so in our original study. We also cannot rule out the potential for different results if a different health outcome were considered.[32]

Previous metrics had not generally incorporated information about phases of well development and none had included compressor stations. Compressor stations were responsible for the majority of UNGD-related emissions of VOC, nitrogen oxides, and PM_{10} and $PM_{2.5}$ (particulate matter less than or equal to 10 and 2.5 μm in aerodynamic diameter, respectively) in Pennsylvania in 2011.[7] Therefore, the inclusion of compressors likely would improve the ability to estimate UNGD-related air pollution with the $IDW_2 4C$ metric. Recently, Allshouse and colleagues recommended the use of an UNGD activity index that incorporated the four phases of UNGD and the number of gas storage tanks at the well site.[52] They found a Spearman correlation between their metric and air pollutants measured at 25 locations of 0.74; however, they did not include flaring or impoundments. Estimates of the contribution of flaring to UNGD air emissions range from <0.1% of VOCs to an increase of 120% with flaring.[7,53] Impoundments remain a largely uncharacterized source of air emissions.[16]

One important pathway through which UNGD could influence health is air quality impacts. There are other secondary data sources that could be considered in evaluation of these impacts. We wanted to explicitly evaluate the adequacy of a GIS-based metric for air quality impacts by comparing UNGD activity metrics to air quality estimates. To do this, we needed air pollution measurements that were on a fine spatial and temporal resolution that included emissions from UNGD and covered the years of UNGD (2005−2015) in Pennsylvania. Because EPA monitors were prohibitively sparse in counties with UNGD (Figure 1) to conduct kriging analysis, we considered using the community multiscale air quality (CMAQ) model output on a 12-km grid for $PM_{2.5}$ and ozone in 2007 and 2011. However, the national emissions inventory (NEI), which CMAQ uses, "likely underestimates oil and gas emissions."[54] It included only 2675 unconventional natural gas wells in Pennsylvania in 2011, whereas our analysis identified 4951 spudded wells by the end of 2011. The EPA is working to improve UNGD emissions for future versions of the NEI, so it may be possible to validate UNGD activity metrics to a surrogate for air quality impacts using CMAQ in the future.

This study had additional limitations. We likely underestimated the number of compressor engines because we could not distinguish between compressor engines missing a start letter and those never started. Like UGD wells,[34] some compressors were likely planned, but never put in place. Additionally, we were not able to evaluate if the original list of UNG-related compressor stations from the DEP was missing any stations. We did not have information on whether compressor engines were diesel or natural gas powered. We also restricted the PCA to the Geisinger region, and therefore it may not be generalizable to areas where wells, compressors, and impoundments are not co-located or other shale formations with different well development and production practices. Finally, while we considered compressors, impoundments, and flaring in our new exposure model, future studies should consider incorporating information on fugitive emissions, pumps, and tanks, which may represent disproportionate sources of VOC and methane emissions related to UNGD.

DOI: 10.1021/acs.est.8b00507
Environ. Sci. Technol. 2018, 52, 6061−6069

No prior study has described the size and temporal and spatial distribution of UNGD-related compressor stations, impoundments, and flaring in Pennsylvania, and evaluated what information they added to GIS-based metrics. We also compared associations of previously used UNGD metrics and a newly developed $IDW_2$4C metric with mild asthma exacerbations. We found that GIS proxies for UNGD were defensible metrics to retrospectively capture multiple pathways for low cost in the initial studies of UNGD and health. Given the small potential benefits of including flaring and impoundment data in UNGD activity metrics, it does not seem to be worth the time and effort required to obtain the data. Based on these findings, we recommend that researchers use the $IDW_2$4C metric, when possible. Even when relying on secondary data over large geographies, researchers can extract additional information about UNG-related exposures by incorporating time-varying phase-specific well features. For example, the volume of production is likely a good surrogate for compressor activity and well depth a surrogate for truck trips and air quality impacts associated with well completion (i.e., hydraulic fracturing). We acknowledge that without environmental measurements, it is not possible to determine what pathways are captured by the GIS proxies, and this study highlights the need for future UNGD and health studies to improve exposure assessment by collecting environmental measurements or biomarkers. Only when we understand how UNGD is affecting health can we effectively design interventions to reduce exposure. In the face of resource constraints, however, public health researchers can implement the $IDW_2$4C metric as a likely proxy for multiple exposure pathways related to UNGD.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.est.8b00507.

Additional principal component results, and impoundment and flaring data access (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: bschwar1@jhu.edu.

ORCID

Kirsten Koehler: 0000-0002-0516-6945

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This study was funded by the National Institute of Environmental Health Sciences Grant ES023675-01 (PI: B.S.S), training Grant ES07141 (S.G. Rasmussen), and Grant K99ES027023 (J.A.C.); the Degenstein Foundation; and the National Science Foundation Integrative Graduate Education and Research Traineeship (S.G. Rasmussen). No funders had input into the study design, conduct, data collection or analysis, or manuscript preparation. Dr. Schwartz 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. We thank Sara Rasmussen for her contributions to the work (Johns Hopkins Bloomberg School of Public Health); Joseph J. DeWalle (Geisinger Health System) for creating the map; John Amos (SkyTruth) for guidance with the flaring and impoundment data; and Chloe Quinlan (Johns Hopkins Bloomberg School of Public Health), Jennifer Irving, and Joshua Crisp (Geisinger Health System) for compiling the compressor data.

■ REFERENCES

(1) U.S. Energy Information Administration Shale in the United States. https://www.eia.gov/energy_in_brief/article/shale_in_the_united_states.cfm#shaledata (accessed Feb 10, 2017).

(2) Adgate, J. L.; Goldstein, B. D.; McKenzie, L. M. Potential public health hazards, exposures and health effects from unconventional natural gas development. *Environ. Sci. Technol.* **2014**, *48* (15), 8307–20.

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

(4) Gopalakrishnan, S.; Klaiber, H. A. Is the shale energy boom a bust for nearby residents? Evidence from housing values in Pennsylvania. *Am. J. Agric Econ* **2014**, *96* (1), 43–66.

(5) 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* (0), 203–9.

(6) Jackson, R. B.; Vengosh, A.; Darrah, T. H.; Warner, N. R.; Down, A.; Poreda, R. J.; Osborn, S. G.; Zhao, K.; Karr, J. D. Increased stray gas abundance in a subset of drinking water wells near Marcellus shale gas extraction. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (28), 11250–5.

(7) Litovitz, A.; Curtright, A.; Abramzon, S.; Burger, N.; Samaras, C. Estimation of regional air-quality damages from Marcellus Shale natural gas extraction in Pennsylvania. *Environ. Res. Lett.* **2013**, *8* (1), 014017.

(8) Maloney, K. O.; Yoxtheimer, D. A. Production and disposal of waste materials from gas and oil extraction from the Marcellus Shale play in Pennsylvania. *Environ. Pract.* **2012**, *14* (4), 278–287.

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

(10) Muehlenbachs, L.; Spiller, E.; Timmins, C. *The Housing Market Impacts of Shale Gas Development*; National Bureau of Economic Research, 2014.

(11) Multi-state Shale Research Collaborative The Shale Tipping Point: The Relationship of Drilling to Crime, Traffic Fatalities, STDs, and Rents in Pennsylvania, West Virginia, and Ohio. http://www.multistateshale.org/shale-tipping-point (accessed Dec 22, 2017).

(12) Olmstead, S. M.; Muehlenbachs, L. A.; Shih, J.-S.; Chu, Z.; Krupnick, A. J. Shale gas development impacts on surface water quality in Pennsylvania. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (13), 4962–4967.

(13) Osborn, S. G.; Vengosh, A.; Warner, N. R.; Jackson, R. B. Methane contamination of drinking water accompanying gas-well drilling and hydraulic fracturing. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (20), 8172–6.

(14) Pacsi, A. P.; Alhajeri, N. S.; Zavala-Araiza, D.; Webster, M. D.; Allen, D. T. Regional air quality impacts of increased natural gas production and use in Texas. *Environ. Sci. Technol.* **2013**, *47* (7), 3521–3527.

(15) Pacsi, A. P.; Kimura, Y.; McGaughey, G.; McDonald-Buller, E. C.; Allen, D. T. Regional ozone impacts of increased natural gas use in the Texas power sector and development in the Eagle Ford shale. *Environ. Sci. Technol.* **2015**, *49* (6), 3966–3973.

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

DOI: 10.1021/acs.est.8b00507
Environ. Sci. Technol. 2018, 52, 6061–6069

(17) Sangaramoorthy, T.; Jamison, A. M.; Boyle, M. D.; Payne-Sturges, D. C.; Sapkota, A.; Milton, D. K.; Wilson, S. M. Place-based perceptions of the impacts of fracking along the Marcellus Shale. *Soc. Sci. Med.* **2016**, *151*, 27−37.

(18) Vengosh, A.; Jackson, R. B.; Warner, N.; Darrah, T. H.; Kondash, A. A critical review of the risks to water resources from unconventional shale gas development and hydraulic fracturing in the United States. *Environ. Sci. Technol.* **2014**, *48* (15), 8334−48.

(19) Vinciguerra, T.; Yao, S.; Dadzie, J.; Chittams, A.; Deskins, T.; Ehrman, S.; Dickerson, R. R. Regional air quality impacts of hydraulic fracturing and shale natural gas activity: Evidence from ambient VOC observations. *Atmos. Environ.* **2015**, *110* (0), 144−150.

(20) Warner, N. R.; Jackson, R. B.; Darrah, T. H.; Osborn, S. G.; Down, A.; Zhao, K.; White, A.; Vengosh, A. Geochemical evidence for possible natural migration of Marcellus Formation brine to shallow aquifers in Pennsylvania. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (30), 11961−6.

(21) Brasier, K. J.; Rhubart, D. Effects of Marcellus shale development on the criminal justice system (The Marcellus impacts Project Report # 6). http://www.rural.palegislature.us/documents/reports/Marcellus-Report-6-Crime%20.pdf, 2017.

(22) Casey, J. A.; Savitz, D. A.; Rasmussen, S. G.; Ogburn, E. L.; Pollak, J.; Mercer, D. G.; Schwartz, B. S. Unconventional natural gas development and birth outcomes in Pennsylvania, USA. *Epidemiology* **2016**, *27* (2), 163−172.

(23) Currie, J.; Greenstone, M.; Meckel, K. Hydraulic fracturing and infant health: New evidence from Pennsylvania. *Sci. Adv.* **2017**, *3* (12), e1603021.

(24) 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* (4), 412−417.

(25) 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* (6), e0126425.

(26) Saberi, P.; Propert, K. J.; Powers, M.; Emmett, E.; Green-McKenzie, J. Field survey of health perception and complaints of Pennsylvania residents in the Marcellus Shale region. *Int. J. Environ. Res. Public Health* **2014**, *11* (6), 6517−27.

(27) 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* (2), 189−197.

(28) 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* **2014**, *123* (1), 21−26.

(29) Finkel, M. L. Shale gas development and cancer incidence in southwest Pennsylvania. *Public Health* **2016**, *141*, 198−206.

(30) Fryzek, J.; Pastula, S.; Jiang, X.; Garabrant, D. H. Childhood cancer incidence in Pennsylvania counties in relation to living in counties with hydraulic fracturing sites. *J. Occup. Environ. Med.* **2013**, *55* (7), 796−801.

(31) Jemielita, T.; Gerton, G. L.; Neidell, M.; Chillrud, S.; Yan, B.; Stute, M.; Howarth, M.; Saberi, P.; Fausti, N.; Penning, T. M.; et al. Unconventional gas and oil drilling is associated with increased hospital utilization rates. *PLoS One* **2015**, *10* (7), e0131093.

(32) 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* (9), 1334−43.

(33) Pennsylvania Department of Conservation and Natural Resources Pennsylvania Internet Record Imaging System Well Information System. http://www.dcnr.state.pa.us/topogeo/econresource/oilandgas/resrefs/wis_home/ (accessed Dec 26, 2017).

(34) Pennsylvania Department of Environmental Protection PA DEP Oil & Gas Reporting Website. https://www.paoilandgasreporting.state.pa.us/publicreports/Modules/Welcome/Welcome.aspx (accessed Dec 26, 2017).

(35) United States Department of Agriculture (USDA) National Agriculture Imagery. https://www.fsa.usda.gov/programs-and-services/aerial-photography/imagery-programs/naip-imagery/index (accessed Dec 26, 2016).

(36) Verbesselt, J.; Hyndman, R.; Newnham, G.; Culvenor, D. Detecting trend and seasonal changes in satellite image time series. *Remote Sens Environ* **2010**, *114* (1), 106−115.

(37) Platt, R. V.; Manthos, D.; Amos, J. Estimating the creation and removal date of fracking ponds using trend analysis of Landsat imagery. *Environ. Manage.* **2018**, *61*, 310.

(38) Lever, J.; Krzywinski, M.; Altman, N. POINTS OF SIGNIFICANCE Principal component analysis. *Nat. Methods* **2017**, *14* (7), 641−642.

(39) Ogden, C. L.; Carroll, M. D.; Kit, B. K.; Flegal, K. M. Prevalence of childhood and adult obesity in the United States, 2011−2012. *JAMA* **2014**, *311* (8), 806−14.

(40) Casey, J. A.; Pollak, J.; Glymour, M. M.; Mayeda, E. R.; Hirsch, A. G.; Schwartz, B. S. Measures of SES for Electronic Health Record-based Research. *Am. J. Prev. Med.* **2018**, *54* (3), 430−439.

(41) Schwartz, B. S.; Stewart, W. F.; Godby, S.; Pollak, J.; DeWalle, J.; Larson, S.; Mercer, D. G.; Glass, T. A. Body mass index and the built and social environments in children and adolescents using electronic health records. *Am. J. Prev. Med.* **2011**, *41* (4), e17−e28.

(42) Jackson, J. E. *A User's Guide to Principal Components*; John Wiley & Sons, 2005; Vol. *587*.

(43) Legendre, P. L. L. *Numerical Ecology*, 3rd ed.; Elsevier: London, UK, 2012.

(44) Abualfaraj, N.; Olson, M. S.; Gurian, P. L.; De Roos, A.; Gross-Davis, C. A. Statistical analysis of compliance violations for natural gas wells in Pennsylvania. *Energy Policy* **2016**, *97*, 421−428.

(45) Skytruth Fracking Chemical Database. http://frack.skytruth.org/fracking-chemical-database (accessed Nov 13, 2017).

(46) Schmid, K. W. The Marcellus Shale Gas Play-Geology and Production and Water Management, Oh My! *Pennsylvania Geology* **2012**, *42* (2), 3−12.

(47) Lane, K. J.; Levy, J. I.; Scammell, M. K.; Patton, A. P.; Durant, J. L.; Mwamburi, M.; Zamore, W.; Brugge, D. Effect of time-activity adjustment on exposure assessment for traffic-related ultrafine particles. *J. Exposure Sci. Environ. Epidemiol.* **2015**, *25* (5), 506−16.

(48) Setton, E.; Marshall, J. D.; Brauer, M.; Lundquist, K. R.; Hystad, P.; Keller, P.; Cloutier-Fisher, D. The impact of daily mobility on exposure to traffic-related air pollution and health effect estimates. *J. Exposure Sci. Environ. Epidemiol.* **2011**, *21* (1), 42−8.

(49) Adgate, J. L.; Church, T. R.; Ryan, A. D.; Ramachandran, G.; Fredrickson, A. L.; Stock, T. H.; Morandi, M. T.; Sexton, K. Outdoor, indoor, and personal exposure to VOCs in children. *Environ. Health Perspect* **2004**, *112* (14), 1386−92.

(50) Jung, K. H.; Artigas, F.; Shin, J. Y. Personal, indoor, and outdoor exposure to VOCs in the immediate vicinity of a local airport. *Environ. Monit. Assess.* **2011**, *173* (1−4), 555−67.

(51) Nethery, E.; Leckie, S. E.; Teschke, K.; Brauer, M. From measures to models: an evaluation of air pollution exposure assessment for epidemiological studies of pregnant women. *Occup. Environ. Med.* **2008**, *65* (9), 579−86.

(52) Allshouse, W. B.; Adgate, J. L.; Blair, B. D.; McKenzie, L. M. Spatiotemporal industrial activity model for estimating the intensity of oil and gas operations in Colorado. *Environ. Sci. Technol.* **2017**, *51* (17), 10243−10250.

(53) New York State Department of Environmental Conservation. *Revised Draft Supplemental Generic Environmental Impact Statement (EIS) on the Oil, Gas and Solution Mining Regulatory Program*; 2011.

(54) U.S. Environmental Protection Agency. *EPA Needs to Improve Air Emissions Data for the Oil and Natural Gas Production Sector*; Washington, D.C, 2013.

DOI: 10.1021/acs.est.8b00507
Environ. Sci. Technol. 2018, 52, 6061−6069