FINAL REPORT: CONTRACT CARB No. 18RD018

Characterizing the Potential Health and Equity Impacts of Oil and Gas Extraction and Production Activities in California

Principal Investigator:
Rachel Morello-Frosch, PhD, MPH
Department of Environmental Science Policy and Management & School of Public Health,
University of California, Berkeley

Prepared for the California Air Resources Board (CARB)

Rachel Morello-Frosch, PhD, MPH
Department of Environmental Science, Policy and Management & School of Public Health,
University of California, Berkeley

Kathy Tran, PhD, MPH
School of Public Health, Environmental Health Sciences Division
University of California, Berkeley

Lauren Baehner, MPH
Department of Environmental Science, Policy and Management & School of Public Health,
University of California, Berkeley

Contact: Rachel Morello-Frosch, PhD, MPH
University of California, Berkeley
Department of Environmental Science, Policy and Management
& School of Public Health
130 Mulford Hall, Berkeley, 94720-3114
Email: rmf@berkeley.edu
Work Phone: 510-643-6358

Disclaimer
The statements and conclusions in this report are those of the researchers on this contract and not necessarily those of the California Air Resources Board. The mention of commercial products, their source, or their use in connection with material reported herein is not to be construed as actual or implied endorsement of such products.

Acknowledgements
The authors express their gratitude to the many individuals involved in supporting and/or providing valuable feedback on this work, including Nicholas Depsky, Jessie Jaeger, Sylvia Sudat, Seigi Karasaki, Clare Pace, Yang Ju, David Gonzalez, Seth Shonkoff, as well as research scientists and staff from CARB, OEHHA and the State Water Resources Control Board. Finally, this project would not have been possible without the dedication of and outstanding feedback from community organizations across California whose tireless work to advance environmental justice and ensure that state policies and regulations that address oil and gas development protect public health for everyone. In addition to primary support for this project from the California Air Resources Board (# 18RD018), funding from the National Institutes of Environmental Health Sciences (R00 ES027023 and P30 ES009089) helped support Dr. Joan Casey's work on this project. All errors are the responsibility of the authors.

TABLE OF CONTENTS

List of Tables

List of Figures

Abstract
There is a paucity of studies in California that have assessed the impacts of exposure to oil and gas development (OGD) on perinatal health outcomes as well as potential threats to drinking water sources, particularly among communities reliant upon domestic wells. Similarly, high methane emitters, which include oil and gas production and distribution sites as well as landfills, dairies, and refineries, can emit non-methane co-pollutants that are harmful to human health; yet this category of climate change hazard has been understudied in terms of its implications for environmental justice and potential acute health effects.

For this project, we examined the relationship between perinatal health outcomes and: 1) exposure to active and inactive wells, accounting for production volume during trimester of pregnancy, and 2) hydraulic fracturing (HF) during pregnancy. We also conducted a spatial analysis of OGD infrastructure sites and domestic wells areas (DWA-- populated areas served by at least one domestic well) and community water systems (CWS-- public drinking water systems with at least 15 connections) to identify potential groundwater threats to then determine whether at-risk drinking water sources in the San Joaquin Valley (SJV) serve vulnerable populations. Finally, we examined the relationship between proximity to high methane emitters and migraine prevalence and exacerbation and conducted an equity assessment of community proximity to and exposure intensity of California's high methane emitters. Although not directly toxic to humans, methane is co-emitted with other harmful pollutants that do threaten the health of nearby communities.

Results from our epidemiological studies on the perinatal effects of proximity to OGD and HF showed positive associations between these exposures and adverse birth outcomes, including increased odds of small for gestational age and low birth weight births as well as decreased term birth weight. Overall, effect estimates were stronger among births in rural compared to urban areas. Our spatial analysis of potential threats of OGD infrastructure sites showed that CWS intersecting OGD infrastructure had fewer residents within the system per km^2, a lower proportion of residents living twice below poverty and higher proportion of Latinos compared to CWS that did not intersect OGD. Models showed that small CWS (less than 15 connections) significantly predicted higher counts of OGD infrastructure compared to larger systems.

Analyses of locations of high methane emitters, emissions, and levels of other pollutants showed increased odds of migraine case status with increasing methane emissions. Results also showed increased odds of migraine case status with higher NO_2 levels. We found no association between $PM_{2.5}$ levels or proximity to oil and gas wells and migraine case status. $PM_{2.5}$ and NO_2 were positively associated with migraine exacerbation outcomes and we observed limited or null associations between continuous measures of methane emissions and proximity to oil and gas wells and migraine severity. In our equity assessment of high methane emitters in California, we observed environmental injustice in the locations of high methane emitters and emissions intensity for those block groups with higher proportions of residents of color and lower voter turnout. We did not observe associations with measures of socioeconomic status. Some of these significant associations were non-linear.

Results from these analyses indicate the importance of holistically characterizing the potential human health and equity implications of OGD as well as other climate change hazards, including

high methane emitters to ensure that regulatory decision-making for these sites integrates public health, sustainability and environmental justice goals. Future studies on the health effects of OGD in California should better characterize the diverse exposures associated with these activities (e.g., air and water contamination, noise, excessive light, and other stressors). Moreover, additional health outcomes should be studied, including respiratory, cardiovascular, and developmental outcomes. Future research on high methane emitters would benefit from more consistent temporal monitoring to assess changes in emission trends and to better characterize methane and co-pollutant emission relationships. Modeling approaches can also estimate acute and chronic exposures to potentially harmful co-pollutants from high methane emitters and support additional studies on their health effects on communities living nearby.

List of Abbreviations

BMI- body mass index
BOE – barrels of oil equivalent
Cal-GEM - California Geologic Energy Management Division—Cal-GEM
CWS – community water systems
DWA – domestic well area
DOGGR - Division of Oil, Gas and Geothermal Resources
ED- emergency department
EHR- electronic health records
HF – hydraulic fracturing
ICE – index of concentration at the extremes
IDW- inverse distance weighted
LBW – low birth weight
MPA- migraine probability algorithm
NO_2- nitrogen dioxide
OGD – oil and gas development
$PM_{2.5}$ - fine particulate matter
PTB- preterm birth
SGA – small for gestational age
SJV- San Joaquin Valley
SO_2 - sulfur dioxide
tBW – term birth weight
VOCs - volatile organic compounds

Project Background and Scientific Executive Summary

As California seeks to address the health and climate change impacts of oil and gas development and other major sources of greenhouse gas emissions, including large methane emitters [also known as methane super-emitters (Duren et al. 2019)], there is a paucity of studies to inform regulatory decision-making on these categories of environmental hazards. Domestic oil and gas development activities in California and nationally, including unconventional extraction methods (also referred to as well stimulation) that include horizontal drilling and the intensive use of chemicals to release oil and gas from the ground, have raised concerns about potential adverse health impacts on local communities due to increases in air pollution, noise, and water contamination, among other factors. This environmental health issue has also gained traction in the regulatory arena due to the ubiquity of oil and gas development sites in California, many of which are near sensitive receptors. California has large oil reserves with active and inactive wells that are located near densely populated and rural areas primarily in the San Joaquin Valley and Los Angeles Air Basins. Los Angeles is unique in that oil and gas production developed simultaneously with the growth of the city. Residential proximity to oil and gas activities may increase exposures to air pollutant emissions and other results of oil and gas development activities (e.g., water use, dust, chemicals, excessive noise and light). Households that use groundwater from private drinking water wells or small community drinking water systems that rely on groundwater sources located near oil and gas development may be at increased risk of drinking water contamination. Previous studies in Pennsylvania, Texas and Colorado have found adverse birth outcomes associated with proximity to unconventional natural gas development (UNDG) activities. However, the process of oil and gas development and extraction in California is different; the state's oil and gas infrastructure is located in both rural and densely populated urban areas. There is a paucity of health studies in California, and, to our knowledge, no analyses have examined the birth outcome effects of oil and gas activities and development in the state, although several studies have linked poor fetal growth outcomes and pre-term birth with various air pollutants (such as PM, ozone, sulfur dioxide, air toxics) and traffic-related air pollution.

While studies have found human health risks attributable to emissions of petroleum related compounds associated with oil and gas development in general, to our knowledge, the public health impacts associated with proximity to upstream oil and gas activities have not been extensively studied in California. Moreover, CARB has identified high methane emitters including oil and gas production sites, landfills, dairies, refineries and other sites. As such, more detailed analyses of the relationship between high methane emitters and acute health effects as well as equity implications of these sites are warranted.

This project proposed to conduct integrated analyses that would characterize the health and environmental equity impacts among vulnerable populations of oil and gas development and activities and high methane emitters in California. The goal was to leverage and integrate a database developed by CARB based on Next Generation Airborne Visible/Infrared Imaging Spectrometer (AVIRIS-NG) flights conducted between 2016–2018 on the location of high methane emitters and their emissions from multiple sources; the Division of Oil, Gas and Geothermal Resources (DOGGR) [now California Geologic Energy Management Division—Cal-GEM] database on active and inactive oil and gas wells and other infrastructure sites; birth records from the California Department of Public Health (2006-2015); and a database developed by this study team on the location and extent of community water systems and domestic well communities in California. Accordingly, this project undertook the following aims:

Aim 1: Assess the association between proximity to all oil and gas development sites (active and inactive wells) during pregnancy and adverse birth outcomes in California.
Aim 2: Evaluate the association of prenatal exposure to hydraulic fracturing (HF) and adverse birth outcomes in urban and rural communities in 8 California counties where HF is prevalent.
Aim 3: Evaluate potential drinking water threats posed by oil and gas sites in the San Joaquin Valley.
Aim 4: Characterize the association between long-term exposure to high methane emitters and other sources of harmful emissions and common air pollutants with both migraine headache and, among patients with migraine, headache severity among patients with migraine; and
Aim 5: Conduct an equity assessment of community proximity and exposure intensity of California's high methane emitters.

The five chapters below are organized by study aim. Four of these analyses have been published or are forthcoming, and one is under review. Citations for those forthcoming and published analyses are included in their corresponding chapter.

Results from our epidemiological studies on the perinatal effects of proximity to OGD and HF showed positive associations between these exposures and adverse birth outcomes, including odds of small for gestational age and low birth weight births as well as decreased term birth weight. Overall, effect estimates were stronger among births in rural compared to urban areas. Our spatial analysis of potential threats of OGD infrastructure sites showed that CWS intersecting OGD infrastructure had fewer residents within the system per km2, a lower proportion of residents living twice below poverty and higher proportion of Latinos compared to CWS that did not intersect OGD. Models showed that small CWS (less than 15 connections) significantly predicted higher counts of OGD infrastructure compared to larger systems.

Analyses of high methane emitters and emissions, as well as concentrations of other air pollutants showed increased odds of migraine case status with increasing methane emissions. Although not directly toxic to humans, methane is co-emitted with other harmful pollutants that do threaten the health of nearby communities. Results also showed increased odds of migraine case status with increasing NO_2 levels. We found no association between $PM_{2.5}$ levels or proximity to oil and gas wells and migraine case status. Levels of $PM_{2.5}$ and NO_2 were positively associated with migraine exacerbation outcomes and we observed limited or null associations between continuous measures of methane emissions and proximity to oil and gas wells and migraine severity. In our equity assessment of high methane emitters in California, we observed environmental injustice in the locations of high methane emitters and emissions intensity for those block groups with higher proportions of residents of color and lower voter turnout. We did not observe associations with measures of socioeconomic status. Some of these significant associations were non-linear.

Results from these analyses indicate the importance of holistically characterizing the potential human health and equity implications of oil and gas development as well as other climate change hazards, including high methane emitters to ensure that regulatory decision-making integrates public health, sustainability and environmental justice goals. Future studies on the health effects of OGD in California should better characterize the diverse exposures associated with these activities (e.g., air and water contamination, noise, excessive light, and other stressors). Moreover, additional health outcomes should be studied, including respiratory, cardiovascular, and developmental outcomes. Future research on high methane emitters would benefit from more consistent temporal monitoring to assess changes in emission trends and to better characterize methane and co-pollutant emission relationships. Modeling approaches can also estimate acute and chronic exposures to potentially harmful co-pollutants from high methane emitters and support additional studies on their health effects on communities living nearby.

Background and Lay Executive Summary

As California works to reduce the health and climate change impacts of oil and gas development and other major sources of greenhouse gas emissions, including large sources of methane emissions [also known as methane super-emitters (Duren et al. 2019)], there are few California studies to make decisions about how best to regulate these types of environmental hazards. Oil and gas development activities in California and across the United States have raised concerns about potential adverse health impacts on local communities due to increases in air pollution, noise, and water contamination, among other factors. These activities include unconventional extraction methods, such as hydraulic fracturing or "fracking" (also referred to as well stimulation) that include horizontal drilling and the large-scale use of chemicals to release oil and gas from the ground.

This environmental health issue has also gained attention in the regulatory arena due to the large number of oil and gas development sites in California, many of which are near communities and sensitive land uses (such as schools, parks and elderly housing) that are disproportionately impacted. California has large oil reserves with active and inactive wells that are located near densely populated and rural areas primarily in the San Joaquin Valley and Los Angeles Air Basins. Los Angeles is unique in that oil and gas production developed at the same time as the growth of the city. Living near oil and gas activities may increase exposures to air pollutant emissions and other hazards associated with oil and gas development activities (e.g., water use, dust, chemicals, excessive noise, and excessive light). Households that rely on private drinking water wells or small community drinking water systems that rely on groundwater sources located near oil and gas development may be at increased risk of drinking water contamination. Previous studies in Pennsylvania, Texas and Colorado have found adverse birth outcomes associated with living near unconventional natural gas development (UNDG) activities. However, the process of oil and gas development and extraction in California is different; the state's oil and gas infrastructure is located in both rural and densely populated urban areas. There is a lack of health studies in California, and, to our knowledge, no analyses have examined the impacts of oil and gas development on birth outcomes, although several studies have linked poor fetal growth outcomes and pre-term birth with various air pollutants (such as particulate matter (PM), ozone, sulfur dioxide, air toxics) and traffic-related air pollution.

While studies have found human health risks associated with emissions of petroleum related compounds associated with oil and gas development in general, to our knowledge, the public health impacts associated with living near upstream oil and gas activities have not been widely studied in California. Moreover, CARB has identified high methane emitters including oil and gas production sites, landfills, dairies, refineries, and other sites. Therefore, more detailed analyses of the relationship between high methane emitters and acute health effects as well as equity implications of these sites are needed.

This project proposed studies that would characterize the health and environmental equity impacts among vulnerable populations of oil and gas development and high methane emitters in California. The goal was to utilize a database developed by CARB on the location of high methane emitters and their emissions from multiple sources; the Division of Oil, Gas and Geothermal Resources (DOGGR) [now California Geologic Energy Management Division—

Cal-GEM] database on active and inactive oil and gas wells and other infrastructure sites; birth records from the California Department of Public Health (2006-2015); and a database developed by this study team on the location and extent of community water systems and domestic well communities in California. Accordingly, this project undertook the following aims:

Aim 1: Examine the association between living near all oil and gas development sites (active and inactive wells) during pregnancy and adverse birth outcomes in California.
Aim 2: Evaluate the association between prenatal exposure to hydraulic fracturing (HF) and adverse birth outcomes in urban and rural communities in eight California counties where HF is widespread.
Aim 3: Evaluate potential drinking water threats posed by oil and gas sites in the San Joaquin Valley.
Aim 4: Analyze the relationship between long-term exposure to high methane emitters and other sources of harmful emissions and common air pollutants with migraine headache and, among those patients with migraine, headache severity; and
Aim 5: Conduct an equity assessment of community proximity and exposure intensity of California's high methane emitters.

The five chapters below are organized by study aim. Four of these analyses have been published or are forthcoming, and one is under review. Citations for those forthcoming and published analyses are included in their corresponding chapter.

Results from our epidemiological studies on the perinatal effects of living near OGD and HF showed that mothers living closer to ODG and HF were more likely to experience adverse birth outcomes, including odds of small for gestational age and low birth weight births as well as decreased term birth weight. Overall, the relationship between living near OGD and HF was stronger for births in rural areas than births in urban areas. Our spatial analysis of potential threats of OGD infrastructure sites showed that CWS intersecting OGD infrastructure had fewer residents within the system per km2, a lower proportion of residents living twice below poverty and higher proportion of Latinos, compared to CWS that did not intersect OGD. Models showed that small CWS (less than 15 connections) significantly predicted higher counts of OGD infrastructure compared to larger systems.

Analyses of high methane emitters and emissions as well as concentrations of other air pollutants showed that increased methane emissions increased the chance of migraine case status Although not directly toxic to humans, methane is co-emitted with other harmful pollutants that do threaten the health of nearby communities. Results also showed increased chance of migraine case status with increasing NO_2 levels. We found no association between $PM_{2.5}$ levels or proximity to oil and gas wells and migraine case status. Levels of $PM_{2.5}$ and NO_2 were positively associated with migraine exacerbation outcomes, and we observed limited or no associations between continuous measures of methane emissions and proximity to oil and gas wells and migraine severity. In our equity assessment of high methane emitters in California, we observed environmental injustice in the locations of high methane emitters and emissions intensity for those block groups with higher proportions of residents of color and lower voter turnout. We did not observe associations with measures of socioeconomic status. Some of these significant associations were non-linear.

Results from these analyses show the importance of holistically characterizing the potential human health and equity implications of oil and gas development as well as other climate change hazards, including high methane emitters to ensure that regulatory decision-making integrates public health, sustainability, and environmental justice goals. Future studies on the health effects of oil and gas development in California should better characterize the diverse exposures associated with these activities (e.g., air and water contamination, noise, excessive light, and other stressors). Moreover, additional health outcomes should be studied, including respiratory, cardiovascular, and developmental outcomes. Future research on high methane emitters would benefit from more consistent temporal monitoring to assess changes in emission trends and to better understand methane and co-pollutant emission relationships. Modeling techniques can also estimate acute and chronic exposures to potentially harmful co-pollutants from high methane emitters and support additional studies on their health effects on communities living nearby.

Chapter 1: Residential proximity to oil and gas development and birth outcomes in California: a retrospective cohort study of 2006-2015 births

Publication: Tran KV, Casey JA, Cushing LJ, Morello-Frosch R (2020) Residential proximity to oil and gas development and birth outcomes in California: a retrospective cohort study of 2006-2015 births. *Environmental Health Perspectives*. 128(6) June 2020
https://doi.org/10.1289/EHP5842

1.1 Abstract

Background: Studies suggest associations between oil and gas development (OGD) and adverse birth outcomes, but few epidemiologic studies of oil wells or inactive wells exist, and none in California.
Objective: To investigate the relationship between residential proximity to OGD and birth outcomes in California.
Methods: We conducted a retrospective cohort study of 2,918,089 births to mothers living within 10 km of at least one production well between January 1, 2006 and December 31, 2015. We estimated exposure during pregnancy to inactive wells count (no inactive wells, 1 well, 2-5 wells, 6+wells) and production volume from active wells in barrels of oil equivalent (BOE) (no BOE, 1-100 BOE/day, > 100 BOE/day). We used generalized estimating equations to examine associations between overall and trimester-specific OGD exposures and term birth weight (tBW), low birth weight (LBW), preterm birth (PTB), and small for gestational age birth (SGA). We assessed effect modification by urban/rural community type.
Results: Adjusted models showed exposure to active OGD was associated with adverse birth outcomes in rural areas; effect estimates in urban areas were close to null. In rural areas, increasing production volume was associated with stronger adverse effect estimates. High (> 100 BOE/day) versus no production throughout pregnancy was associated with increased odds of LBW (odds ratio [OR] = 1.40, 95% CI: 1.14, 1.71) and SGA (OR = 1.22, 95% CI: 1.02, 1.45), and decreased tBW (mean difference = -36 grams, 95 % CI: -54, -17), but not with PTB (OR = 1.03, 95% CI: 0.91, 1.18).
Conclusion: Proximity to higher production OGD in California was associated with adverse birth outcomes among mothers residing in rural areas. Future studies are needed to confirm our findings in other populations and improve exposure assessment measures.

1.2 Background

Oil and gas development (OGD) by the US petroleum industry spans decades in many states but concern about its potential health and equity impacts did not gain traction among researchers until the recent rapid uptake of hydraulic fracturing (HF) (Finkel and Law 2011; Kovats et al. 2014; Mitka 2012). As of 2017, California (CA) was one of the top five producers of crude oil in the country (US EIA 2018a, 2018b). Four of the ten largest US oil fields are in CA's San Joaquin and Los Angeles Basins (Long et al. 2015a) and unlike newer shale gas plays, most of CA's natural gas is extracted from reservoirs also producing oil (Long et al. 2015c). Given the long history of OGD in CA, stimulation techniques, such as water and steam injection and HF, are primarily used at established sites rather than newly drilled wells. Oil recovered via water flooding and steam injection (conventional enhanced oil recovery methods) accounted for 76% of the state's oil production in 2009 (Long et al. 2015c) while HF, an unconventional stimulation technique, accounted for 20% of CA's oil production in the last decade. Due to types of geological formations, HF practices in CA differ from other states, potentially resulting in differing environmental hazards (Long et al. 2015c). OGD production in CA also occurs in both rural and urban settings compared to other states, such as rural Pennsylvania and Colorado, where many epidemiological studies have been conducted (Casey et al. 2015c; Currie et al. 2017; Hill 2018; McKenzie et al. 2014; Rasmussen SG et al. 2016; Tustin et al. 2017a).

Therefore, an epidemiologic study of the relationship between adverse birth outcomes and OGD in CA, a state with a diverse population and the most annual births of any US state, can provide insights about the potential health impacts of OGD exposure within both rural and urban areas. Characterizing exposures related to OGD poses significant measurement challenges because multiple environmental hazards are associated with different stages of extraction and production. OGD involves the development of oil/gas sites and wells (production and injection for enhanced recovery), transport of materials to and from well sites, drilling, operation of equipment to recover oil/gas, and collection and disposal of chemicals and waste separated from the raw oil and gas (Long et al. 2015a). These activities are associated with diverse environmental hazards including air and water pollutants, noise, odors, excessive lighting, and undesired land use changes (Adgate et al. 2014a; Long et al. 2015a). The application of unconventional techniques presumably enhances the environmental burdens as additional toxic chemicals that are used can potentially be released into air, water, and soil (Adgate et al. 2014a; Long et al. 2015a; Macey et al. 2014; Roy et al. 2014a; Vengosh et al. 2014a).

Air pollutants associated with OGD include particulate matter with an aerodynamic diameter of < 2.5μm ($PM_{2.5}$), diesel PM, nitrogen oxides (NO_x), secondary ozone formation, mercury, and volatile organic compounds (VOCs) like benzene, toluene, ethylbenzene and xylene (BTEX) from truck traffic, drilling, hydraulic fracturing, production and flaring (Allshouse et al. 2019; Brantley et al. 2015a; Colborn et al. 2014; Eapi et al. 2014; Esswein et al. 2014; Franklin et al. 2019; Goetz et al. 2015; Koss et al.; Lan et al. 2015; Macey et al. 2014; Marrero et al. 2016; Maskrey et al. 2016; Mellqvist et al. 2017; Roy et al. 2014b; Warneke et al. 2014). Additionally, fugitive toxic air contaminants can escape at the wellhead (Garcia-Gonzales et al. 2019b; Warneke et al. 2014) that might impact health near points of release. Water contaminants associated with OGD include gas-phase hydrocarbons, chemicals mixed in drilling fluids, and naturally occurring salts, metals and radioactive elements within shale that surface with wastewater along with recovered oil and gas and can contaminate potable water via leaks and spills or evaporate (Adgate et al. 2014b; Hildenbrand et al. 2015; Long et al. 2015a; Vengosh et al. 2014b). Noise pollution is associated with well pad construction, truck traffic, drilling, pumps, flaring of gases, and other processes (Allshouse et al. 2019; Blair et al. 2018; Ebisu and Bell 2012; US BLM 2006). Drilling and production activities occur both during the daytime and nighttime, and light pollution has been previously reported as a nuisance in communities undergoing unconventional OGD (Long et al. 2015a), suggesting OGD may impact the health of nearby communities via increased psychosocial stress.

Several OGD-related environmental exposures have been linked to reduced birth weight and gestational age: air pollution e.g., $PM_{2.5}$, NOx, SOx (Basu et al. 2014; Dadvand et al. 2013, 2014; Ebisu and Bell 2012; Long et al. 2015a; Morello-Frosch et al. 2010; Ponce et al. 2005; Ritz et al. 2007), noise pollution (Arroyo et al. 2016; Gehring et al. 2014), some of the chemical compounds found in OGD wastewater (Long et al. 2015a; Valero de Bernabé et al. 2004), and psychosocial distress (Dominguez et al. 2008; Goldenberg et al. 2008; Rondó et al. 2003; Valero de Bernabé et al. 2004). Previous studies examining the relationship between unconventional OGD and birth outcomes provide suggestive evidence of adverse effects. While study designs vary, most have characterized OGD exposure based on the density and distance of HF shale gas wells near the maternal residence in urban and rural Colorado (McKenzie et al. 2014, 2019), Pennsylvania (Casey et al. 2015c; Currie et al. 2017; Hill 2018; Ma 2016; Stacy et al. 2015), Oklahoma (Janitz et al. 2019), and urban Texas (Walker Whitworth et al. 2018; Whitworth et al. 2017). Among the 10 studies, 8 evaluated our outcomes of interest. Some studies found greater exposure to OGD was associated with reductions in term birth weight (Hill 2018; Stacy et al. 2015) and increased odds or incidence of low birth weight (Currie et al. 2017; Hill 2018), preterm birth (Casey et al. 2015c; Walker Whitworth et al. 2018; Whitworth et al. 2017) and small for gestational age births (Hill 2018; Stacy et al. 2015). However, these studies also reported statistically insignificant (Casey et al. 2015c; Whitworth et al. 2017) or inverse associations (McKenzie et al. 2014; Stacy et al. 2015) for some birth outcomes.

Building on this research, our study focused on OGD in CA. We conducted our analysis in regions where OGD is concentrated: the Sacramento Valley, San Joaquin Valley, South Central Coast and South Coast air basins. To our knowledge, our retrospective cohort study with births from 2006-2015 is the first to evaluate prenatal OGD exposure from oil as well as gas wells, inactive as well as active wells, and non-HF and HF wells in rural and urban settings of CA.

1.3 Methods

Study population

Birth records for January 1, 2006 to December 31, 2015 were obtained from the CA Department of Public Health (CDPH). CDPH collects statewide birth records that include mother's residential address at the time of birth, which we geocoded to assign exposure to OGD exposure and area-level covariates using ArcGIS (ESRI, Redlands, CA). Births with missing street-level addresses or that could not be successfully geocoded after a manual cleaning of the address fields for spelling and punctuation errors were excluded (5%). We selected the Sacramento, San Joaquin Valley, South Central Coast and South Coast air basins because they had the highest well densities in CA between 2005 and 2015 (**Supplemental Figure 1.1**). We illustrate the construction of the study population in **Figure 1.1**. Exclusion criteria included: missing last menstrual period (LMP) date, which was approximated as the date of conception and used to estimate gestational age (3%); congenital anomalies or abnormal birth conditions such as cleft lip and Down's syndrome (4%); plural births e.g. twins, triplets (4%); implausible birth weights of less than 500 grams or greater than 5500 grams (4%) (Alexander et al. 1996; Padula et al. 2014; Ponce et al. 2005; Talge et al. 2014a); and implausible gestational ages of less than 22 or greater than 44 weeks (4%) (Alexander et al. 1996; Talge et al. 2014a). To limit unmeasured confounding and enhance comparability of exposed and unexposed populations, we also excluded births to mothers who did not live within 10 km of at least one oil/gas production well (3%). Finally, we excluded observations with any missing covariates or outcomes (2%) to arrive at a final study population of 2,918,089 births (N=2,718,629 term births). All study protocols were approved by the Institutional Review Board of the CA Department of Public Health (#13-05-1231) and the University of California, Berkeley (# 2013-10-5693).

Birth outcomes

We assessed the relationship between OGD and four outcomes: 1) continuous birth weight (grams) among term (≥37 completed weeks) births (tBW), 2) low birth weight (LBW) (<2500 grams), 3) preterm birth (PTB) (<37 weeks) and 4) small for gestational age birth (SGA) (birth weight less than the US sex-specific 10th percentile of weight for each week of gestation (Talge et al. 2014a). Gestational age was estimated by subtracting the LMP date from the date of birth.

Exposure assessment

Active and inactive oil and gas well records including monthly production data were downloaded from the California Division of Oil, Gas and Geothermal Resources website (CA DOGGR) in December 2015 (the division has been renamed to the CA Geologic Energy Management Division, CalGEM, as of January 2020). We assessed exposure to inactive wells because previous studies have found fugitive methane emissions from abandoned production wells that have not been plugged or improperly plugged (Boothroyd et al. 2016; Kang et al. 2016; US EPA 2018). VOCs, such as BTEX and toxic air contaminants, are likely co-emitted with methane (LACDPH 2018; SCAQMD 2019), and exposure to VOCs, including BTEX and formaldehyde, are associated with adverse birth outcomes (Bolden et al. 2015; Chang et al. 2017; Maroziene and Grazuleviciene 2002). Some of the 224,695 wells in the dataset began producing as far back as 1900. The DOGGR data included well latitude/longitude and monthly production volume (barrels of oil and/or cubic meters of natural gas). We defined a production well as active if it produced at least one unit of oil or gas in a given month; production wells could transition between active and inactive status across the study period. We combined these well data with mothers' residential addresses at the time of delivery, date of conception (defined as LMP), and date of delivery to assign prenatal exposure to oil and gas wells.

Study participants lived within 10 km of at least one active or inactive well at the time of delivery. Exposure metrics were developed based on active and inactive wells within 1 km (**Figure 1.2A-B**); prior literature suggests highest exposure to OGD-related hazards within this radius (Boyle et al. 2017; McKenzie et al. 2012a; Meng 2015a; Walker Whitworth et al. 2018; Whitworth et al. 2017). We selected the 1 km buffer presuming that localized air pollution is likely the greatest contributor to OGD-related exposure in CA. We used the short distance to minimize the impact of dispersion and the contribution of exposure from other sources of air pollution. We calculated exposure across the entire pregnancy and by trimester to examine potential critical windows of prenatal exposure.

Exposure to active wells was characterized by oil and gas production volume during pregnancy and exposure to inactive wells by well count. Total production volume exposure from active wells within 1 km was derived by summing monthly barrels of oil and barrels of oil equivalent (BOE) of natural gas. Production volume from oil and gas wells were summed because 95% of gas wells also produced oil (i.e., wet gas) and gas-only wells did not produce significant amounts of gas. Production volume was summed as shown in Equation 1:

$$Total\ production\ volume_j = \sum_{i=1}^{n}\sum_{k=k}^{l} Prod(oil)_{ik} + \sum_{i=1}^{n}\sum_{k=k}^{l} Prod(gas)_{ik}/6,$$

where $Prod(oil)_{ik}$ was the production volume of oil (in barrels) and $Prod(gas)_{ik}$ the production volume of gas (in thousand cubic feet, mcf) at well *i* during month and year *k* of mother *j*'s entire pregnancy or trimester. *K* is the month and year of conception or beginning of a trimester, and *l* is the month and year of delivery or end of a trimester. *K* has a minimum value of 1 equal to January 2005, and *l* has a maximum of 124 or December 2015. Gas production volume was converted from the original units to BOE by dividing by 6 since 6 thousand cubic feet (mcf) = 1 BOE (Bonavista Energy Corporation 2018; Schmoker and Klett 2005). The total production volume for the first and last month of the entire pregnancy or trimester was also weighted by the proportion of the month the mother was pregnant.

We calculated the number of inactive wells within 1 km of a mother's residence during her pregnancy by subtracting the number of active wells from the total number of wells within 1 km. For analysis, we first normalized production volume by the number of days of the entire pregnancy or within each trimester by dividing production volume by the total number of days and then categorized exposure to production volume of active wells based on the exposure distribution as: 1) no BOE from active wells, 2) 1-100 BOE/day (moderate), 3) more than 100 BOE/day (high). We similarly categorized exposure to inactive wells as: 1) no inactive wells, 2) 1 inactive well, 3) 2-5 inactive wells, 4) 6 or more inactive wells. The production volume was normalized to prevent bias from neonates born later as their exposure period was longer. Given a lack of *a priori* knowledge about the production volume or inactive well count that might constitute a harmful exposure, we selected these categories based on the distribution of each exposure metric across cases and non-cases to ensure sufficient overall sample size and number of cases in each exposure group. The exposure variables were not modeled as continuous because the distribution was right skewed (**Supplemental Figure 1.2**). Both active and inactive well exposure variables were included in all regression models. The exposure variables were generated in R version 3.3.1.

Covariates

Individual-level covariates that were identified *a priori* as significant predictors of our outcomes and potential confounders based on prior studies were derived from the CDPH birth records. Infant covariates included sex, month (categorical) and year of birth (categorical) to control for seasonal and secular trends. Maternal covariates included age in years (<20, 20-24, 25-29, 30-34, 35+), race/ethnicity (non-Hispanic White, Black, American Indian, Asian-Pacific Islander, unknown or other, and Hispanic), educational attainment (<high school, high school graduate/GED, some college, college+), Kotelchuk index of prenatal care (inadequate, intermediate, adequate, adequate+) (Alexander and Kotelchuck 1996; Kotelchuck 1994), and parity (nulliparous vs. multiparous). For maternal race/ethnicity, American Indian, unknown and

other were combined into one category due to the small number of women in each group. We included mean-centered and mean-centered squared variables for gestational age in the tBW model to allow for nonlinearity.

We also integrated area-level variables, including indicators for air basin and census tract-based urban/rural status, modeled nitrogen dioxide (NO_2) concentrations, and a measure of income concentration. These covariates accounted for neighborhood and regional differences in air quality, economic activity, and emission sources (Arruti et al. 2011; Finkelstein et al. 2003; O'Neill et al. 2003; Wunderli and Gehrig 1990; Zhao et al. 2009). We used 2014 air basin boundaries designated by the California Air Resources Board (CARB 2014), which coincide with county boundaries and roughly delineate areas with similar air quality, meteorology, and geography. We used US Census urban areas (defined as a densely developed territory consisting of urbanized areas of 50,000 or more and urbanized clusters with between 2,500 to 50,000 people (US Census Bureau)) to designate census tracts as urban or rural. Using 2010 boundaries, we categorized census tracts as urban if 60% or more of the tract overlapped with an urban area. We assigned, based on LMP year, tract-level annual ambient NO_2 concentration as a proxy for traffic-related air pollution (Kim et al. 2020a). Lastly, we used the Index of Concentration at the Extremes (ICE) for income as a measure of neighborhood relative deprivation or affluence based on household income by census tract (Massey 1996). ICE provides information about concentration of privilege and deprivation of communities and has previously been associated with infant mortality (Krieger et al. 2016). ICE ranges from -1 to 1, where negative values indicate a concentration of household incomes in the lower 20th percentile of area median household income, while positive values indicate a concentration of household incomes in the higher 80th percentile. We calculated ICE using 2006-2010 ACS and 2011-2015 ACS metropolitan area median household income to establish percentile cutoff values that account for regional differences in the cost of living. These values were then used in combination with census tract median household income from the ACS data of the vintage of the birth year to assign a tract-level ICE value to each birth. For tracts that were not within metropolitan areas, county-level household income cutoffs were used. ICE was categorized by quartile and this categorical variable was included in adjusted models.

Statistical analyses

Statistical analyses were conducted in SAS 9.4 (SAS Institute Inc., Cary, NC). All models were adjusted for individual-level and community-level covariates selected *a priori*: neonate sex, gestational age (tBW model only), month and year of birth, maternal age, race/ethnicity, educational attainment, Kotelchuck index, urban indicator, air basin, NO_2 and ICE for income. Generalized estimating equations were used to account for clustering of mothers within census tracts (Hubbard et al. 2010). Observations with any missing covariate were removed from analyses.

Initial analyses assessed exposure across the entire pregnancy and then during each trimester for the entire study population across the four air basins. Statistical significance was assessed at $\alpha=0.05$. Effect modification (EM) of exposure to active wells by urban/rural status (primary), maternal race/ethnicity and air basin (both secondary) was evaluated via stratification. We report the strata-specific effect estimates and confidence intervals derived from this methodology. To test the heterogeneity between strata-specific estimates, we modeled interaction terms to derive

Bonferroni adjusted p-values for two-sample z-tests using model-estimated beta coefficients and variances (Buckley et al. 2017; UCLA: Statistical Consulting Group). These EM p-values indicate whether the strata-specific associations are statistically significantly different from each other or the referent group. Non-Hispanic Whites were used as the referent in heterogeneity tests for the other racial/ethnic groups because higher rates of adverse birth outcomes have been observed among people of color compared to Whites (Bryant et al. 2010; Teitler et al. 2007). Sacramento Valley was the referent in heterogeneity tests for the other air basins because exposure to active wells were limited to rural areas of that basin, where there were also fewer births. For the effect modification analyses with race/ethnicity and air basin, only exposure across the entire pregnancy was evaluated since trimester-specific estimates were similar to those for the entire pregnancy.

We conducted two sensitivity analyses with exposure variables across the entire pregnancy only. Mothers' smoking status during pregnancy and pre-pregnancy body mass index (BMI) were not collected by CDPH in 2006, so we conducted sensitivity analyses with both of these variables in one model for 2007-2015. Only 2% of mothers smoked during pregnancy among our study population within our study period (prevalence of smoking during pregnancy in CA was 2.5% in 2015) (CDPH 2015). Additionally, we considered potential confounding from other industrial sources of air pollution and included a binary variable for exposure to air pollution from other facilities (e.g. refineries, power plants, metal mining facilities) monitored for emissions including air toxics by the CARB (CARB 2017) within 1 km (referred to as TRI facilities). Only about 2% of mothers resided within proximity to TRI facilities during our study period.

We tested for multi-collinearity between all model variables by calculating the variance inflation factors (Schreiber-Gregory 2012), none of which were high (i.e., > 10). To assess residual spatial dependence, we generated semi-variograms of regression residuals plotted against distance between mothers' residential addresses (Le Rest et al. 2013; SAS) (**Supplemental Figure 1.3**). The residuals appeared randomly distributed, suggesting spatial autocorrelation was likely controlled for by the study design and inclusion of spatial covariates (e.g., NO_2) in regression models.

1.4 Results

Our study included 2,918,089 births in CA between January 2006 and December 2015 located in four air basins: Sacramento, San Joaquin Valley, South Central Coast, and South Coast. The overall mean birth weight was 3,327 grams (SD = 528) (**Table 1.1**). Five percent (N=148,100) of births were LBW, 7% (N=199,460) preterm, and 12% SGA (N=337,943). A maximum of 1,189 inactive wells and 441 active wells were located within 1km of mothers' residences during pregnancy. On average, mothers exposed to moderate production volume (1-100 BOE/day) had 89 inactive and 4 active wells within 1 km of their home during pregnancy, while mothers exposed to high production volume (>100 BOE/day) had an average of 160 inactive wells and 32 active wells within a 1 km buffer. The average moderate total production volume from active wells producing oil and gas during pregnancy was 26 BOE/day, and the average high total production volume was 599 BOE/day. Temporal trends of mean annual production volume and annual rates of the binary birth outcomes showed no distinct patterns in either rural or urban areas (Supplemental Figure 1.4A – 1.4B). Plots of temporal trends in mean annual production volume and mean annual term birth weight also did not reveal consistent patterns in either rural or urban areas (**Supplemental Figures 1.4C – 1.4D**). The reference (no BOE) and exposed populations were relatively similar in terms of demographic and socioeconomic factors (**Table 1.1**). Compared to the reference and moderate production volume groups, mothers within the high production volume category were slightly more educated [35% vs. 23.5%, on average, college or more educated], older [22% vs. 17%, on average, aged 35 or more], more often non-Hispanic [53% vs. 42.5%, on average, non-Hispanic races], more likely to have no previous pregnancies [44% vs. 39.5%, on average, nulliparous], and to reside in urban areas [97% vs. 88%, on average], in the South Coast air basin [94% vs. 68.5%, on average] and in areas with greater wealth [31% vs. 26%, on average, in ICE quartile 4]. Finally, babies born to mothers exposed to high production volume weighed on average 2 and 11 grams less than those born to mothers exposed to moderate production volume and reference group, respectively.

Adjusted models generally found no associations between inactive well count and adverse birth outcomes in both rural and urban areas (**Figure 1.3, Supplemental Tables 1.1-1.2**). All statistically significant associations indicated modestly decreased odds of LBW and PTB (0.96-0.97) (**Figures 1.3A-B; Supplemental Table 1.1**) or minimally increased birth weight (4-5 grams) (**Figure 1.3D; Supplemental Table 1.2**) related to increased inactive OGD well exposure. Models based on trimester-specific exposures yielded similar estimates across trimesters for all four birth outcomes (**Supplemental Tables 1.1-1.2**).

For exposures to production volume from active wells in unstratified models, we observed significant associations between production volume and LBW and SGA (**Supplemental Table 1.3**). When we stratified models by the urban indicator, we observed significant effect modification with stronger associations between high production volume and LBW (p-value = 0.01, **Supplemental Table 1.4**) and tBW (p-value = 0.001, **Supplemental Table 1.7**) in rural areas (**Figure 1.4**). Compared to the reference group, the odds ratio (OR) for LBW was 1.11 (95% confidence interval [CI]: 0.97, 1.27) (**Supplemental Table 1.4**) and odds ratio for SGA was 1.07 (95% CI: 0.97, 1.19) (**Supplemental Table 1.6**) with exposure to moderate production volume across the entire pregnancy in rural areas versus odds ratios of 1.04 (95% CI: 1.00, 1.09) and 1.03 (95% CI: 1.00, 1.07), respectively, in urban areas **(Figures 1.4A & C**). Exposure to high production volume was associated with an odds ratio of 1.40 (95% CI: 1.14, 1.71) for LBW and an odds ratio of 1.22 (95% CI: 1.02, 1.45) for SGA in rural areas versus odds ratios of 0.99 (95% CI: 0.95, 1.04) and 1.04 (95% CI: 1.01, 1.07), respectively, in urban areas (**Figure 1.4A & C; Tables 1.A4 & 1.A6**). Exposure to high production volume was also associated with decreased term birth weight (Mean difference = -36 grams, 95% CI: -54, -17) for the rural stratum compared to the urban stratum (Mean difference = 1 gram, 95% CI: -5, 8) (**Figure 1.4D; Supplemental Table 1.7**). For LBW, SGA and tBW, the strength of the associations increased with higher production volume among the rural, but not the urban population. In general, exposure to production volume throughout pregnancy was not associated with PTB within rural or urban populations (**Figure 1.4B; Table 1.A5**). Models based on trimester-specific exposures yielded similar estimates and EM p-values for all birth outcomes (**Tables 1.A4-1.A7**), except the third trimester for PTB, where exposure to moderate production volume was associated with increased odds of PTB (OR = 1.06, 95% CI: 1.02, 1.11) and high production volume was associated with decreased odds of PTB in urban areas (OR = 0.82, 95% CI: 0.77, 0.88) (**Table 1.A5**).

Maternal race/ethnicity (**Tables 1.A8-1.A9**) and air basin (**Tables 1.A10-1.A11**) did not significantly modify associations between exposure to active well production volume and birth outcomes. Heterogeneity tests were only conducted on the rural population because the effect sizes across outcomes were greater than those of the urban population. Nearly all strata-specific effect estimates included the null and all EM p-values from heterogeneity tests were insignificant across all outcomes.

Sensitivity analyses that included 1) pre-pregnancy BMI and smoking during pregnancy for 2007-2015 births (**Table 1.A12**) and 2) exposure to TRI facilities (**Table 1.A13**) did not change effect estimates by more than 10%.

1.5 Discussion

CA's oil and gas development primarily uses conventional drilling and enhancement methods and, to a much lesser degree, hydraulic fracturing. To our knowledge, our study is the first to quantify prenatal exposures to both inactive wells and cumulative oil and gas production volume from active wells in proximity to pregnant women and to evaluate differences in associations by rural versus urban areas in CA. In rural areas, we found that exposure to high production volume was significantly associated with increased odds of LBW and SGA and decreased tBW compared to the non-exposed group. In urban areas, exposure within 1 km of high production volume relative to no exposure was only significantly associated with increased odds of SGA; effect estimates for exposure to moderate production volume in rural and urban areas were all insignificant.

One prior study, by McKenzie et al. (2019), evaluated urban/rural residential status as an effect modifier. Although that study examined birth defects, the authors found significantly increased odds for four congenital heart defects in the medium and highest exposure groups (based on an intensity-adjusted inverse-distance weighted well-count metric) relative to the lowest group in rural areas (McKenzie et al. 2019); no significant associations were observed for birth defects in urban areas. These rural versus urban differences in effect estimates align with the stronger effect estimates we observed in rural areas in CA for LBW and tBW. McKenzie et al. (2019) also discovered a potential additive effect from other sources of air pollution besides OGD in their analysis. Here, we considered residual confounding from TRI facilities within 1 km, but inclusion of this covariate did not change the rural/urban strata-specific effect estimates. Nevertheless, there may be residual confounding from other sources of air or drinking water pollution that we could not account for in our analysis. For example, the ratio of produced water from OGD (which can contain naturally occurring or injected organic/inorganic chemicals, chemicals that are reaction byproducts, and radioactive materials) to oil and gas extracted increases with well age (Veil et al. 2004). Certain chemicals from produced water could evaporate into the air or percolate into groundwater sources depending on disposal methods (Long et al. 2015a). Air and water pollution concentrations could differ regionally based on dispersion and hydrological transport patterns. Additionally, individual factors that we could not measure in our study such as maternal occupation, housing quality, indoor air quality, dependence upon groundwater sources for drinking water, and underlying population sensitivity to OGD-related pollutants may have contributed to observed differences in effect estimates between rural and urban settings. In the air pollution literature, the exposure-response relationship between cardiovascular disease mortality and $PM_{2.5}$ is relatively steep at low levels of exposure but flattens out at higher levels (Pope et al. 2009; Smith and Peel 2010). Such exposure-response relationships could apply to the OGD setting where urban dwellers may be less affected by OGD-specific pollutants because OGD as an emission source contributes a relatively small percentage to ambient air pollution levels in urban areas, which tend have higher pollutant concentrations overall from diverse mobile and stationary sources. Indeed, average NO_2 levels among urban areas within our study were double that of rural areas.

Results from our analysis align with prior studies that observed decreased birth weight associated with maternal exposure to OGD activities (Currie et al. 2017; Hill 2018; Stacy et al. 2015). However, associations between exposure to OGD and LBW and SGA from other studies have been mixed, with increased odds (Stacy et al. 2015) or incidence probability (Currie et al. 2017; Hill 2018) as well as decreased odds (McKenzie et al. 2014) or no associations (Casey et al. 2015c; Whitworth et al. 2017). Although the mechanisms by which OGD may adversely affect birth weight outcomes remain uncertain, air pollution and noise may be possible pathways that affect maternal health during pregnancy. During production, operation of various ancillary equipment (e.g. wellhead compressors, pneumatic devices, separators, and dehydrators) to collect and process oil and gas generate air pollutants (Garcia-Gonzales et al. 2019b). Multiple VOCs have been measured at oil and gas wellheads and off-site including BTEX and formaldehyde. At ambient levels, BTEX and formaldehyde have been linked to significant decreases in birth weight (Bolden et al. 2015; Chang et al. 2017; Maroziene and Grazuleviciene 2002). Flaring also occurs with oil-producing and horizontally drilled wells (Franklin et al. 2019) and can contribute to spikes in $PM_{2.5}$, black carbon and VOCs during production (Allshouse et al. 2019; Franklin et al. 2019). Relative to other phases of OGD, excessive noise is minimized during production (Allshouse et al. 2019; Hays et al. 2017a). However, noise from compressor stations often exceed the World Health Organization's recommended 55 dBA at night (Hays et al. 2017a) and noise above 65 dBA were measured 20% of the time between 7:00 PM and 7:00 AM in one study (Allshouse et al. 2019). Excessive noise can lead to annoyance and impaired sleep quality (Hays et al. 2017a), which have been linked to low birth weight (Abeysena et al. 2010; Owusu et al. 2013) and preterm birth (Li et al. 2017).

Unlike previous studies, we found no significant association between exposure to active wells and PTB except in the third trimester in urban areas where moderate exposure appeared harmful and high exposure protective. Exposure to OGD was associated with modestly decreased odds for PTB (Stacy et al. 2015) and increased odds (Casey et al. 2015c) in Pennsylvania and increased odds in Texas (Walker Whitworth et al. 2018; Whitworth et al. 2017). The two Pennsylvania studies were conducted in different regions of Pennsylvania and among different populations (general for Stacy et al. (2015) and patients served by one healthcare provider for Casey et al. (2015)). The inverse association in the Stacy et al. (2015) analysis was only observed for the second quartile of exposure compared to the lowest quartile while the association increased with greater exposure (quartiled) in the Casey et al. (2015) study. In Texas, the association was only significant with the highest level of exposure within 10 miles (Walker Whitworth et al. 2018) and the first and second trimesters with exposure within half a mile (Whitworth et al. 2017). Associations for PTB appear to vary by level of exposure as well as trimester. We only observed significant associations–increased odds with moderate exposure and decreased odds with high exposure–in urban areas in the third trimester. Previous studies on air pollution and birth outcomes have suggested that the first and third trimesters are critical windows of exposure for LBW and PTB (Ritz and Wilhelm 2008; Woodruff et al. 2009). Additionally, the significant inverse association between high OGD exposure and PTB in urban areas may reflect residual confounding or live birth bias. Other SES characteristics that were not controlled for in our models could have led to underlying differences among urban dwellers or their exposure patterns. Moreover, if more highly exposed or more vulnerable mothers were less likely to become pregnant or more likely to experience fetal loss, a so-called "depletion of susceptibles" could have occurred (Raz et al. 2018), and a seemingly protective effect would then be observed. While we could not evaluate fertility patterns or spontaneous abortion in our analysis, a study in Ecuador observed greater odds of spontaneous abortion among women who lived within 5 km downstream of an oil field compared to those who lived at least 30 km upstream of an oil field (San Sebastian et al. 2002).

The inconsistent results across studies may reflect differences in statistical and exposure assessment methods, study population demographics, and OGD infrastructure. First, to limit unmeasured confounding, our analyses restricted the study population to those individuals living within 10 km of at least one active or inactive well at the time of delivery. Similar to Whitworth et al. (2017), we specified the unexposed group as those pregnancies with some well activity, but no well activity within 1 km. Besides their exposure, the control and exposed groups are likely more similar to each other on other characteristics (e.g., unmeasured socioeconomic factors) than a control group selected from greater distances or other regions. Second, we applied a 1 km buffer for our exposure metric without weighting, i.e., without up-weighting wells at a shorter distance from maternal residences. Previous studies used inverse distance weighting (McKenzie et al. 2014; Stacy et al. 2015) or inverse distance squared weighting (Casey et al. 2015c; Walker Whitworth et al. 2018; Whitworth et al. 2017), but often included wells beyond our 1 km buffer. Inverse distance weighting has been applied in many air pollution studies (de Mesnard 2013). While air pollution may be a large contributor to OGD-related exposure, we did not assume that it is the only OGD-related hazard and within such a short distance (1 km) dispersion patterns of OGD pollutants may be relatively uniform. Therefore, we weighted all wells equally within the 1 km buffer. Third, we examined separate effects of inactive wells and active well production volume, while prior studies have not considered inactive wells separately and often only examined the density of (McKenzie et al. 2014; Stacy et al. 2015; Whitworth et al. 2017) or total production volume from unconventional wells (Casey et al. 2015c; Walker Whitworth et al. 2018). Including both inactive and active wells allowed us to distinguish possible differential effects by well type. Fourth, our CA study population was more racially and ethnically diverse than those in other studies conducted in CO and PA, which may contribute to differences in analytical results. Finally, California's OGD infrastructure is older than in other states and utilizes less hydraulic fracturing compared to OGD in PA, CO and other states where production infrastructure is newly established (Long et al. 2015c). These regional differences in OGD infrastructure may affect the type of hazards associated with them and their implications for maternal health and birth outcomes.

Our study is the first to highlight differences in potential health impacts of exposure to active OGD based on total production volume from both oil and gas wells and inactive wells. We did not, however, directly measure OGD environmental impacts via, for example, air or drinking water monitoring near active or inactive wells. Several OGD-related hazards–air toxics, water pollutants, noise, excessive lighting–may elicit a variety of biological responses, but our exposure measure precluded identification of specific pathways through which OGD may affect birth outcomes. Further, the cumulative exposure-response curve of all the potential hazards and health outcomes may differ than that for each individual hazard separately. For example, living within proximity to oil/gas fields and seeing the active rigs daily might induce stress, worry, and lack of sleep (Ferrar et al. 2013; Hirsch et al. 2018; Long et al. 2015a; Palagini et al. 2014). However, individuals may habituate, leading to biological responses that may peak and level off (Basner et al. 2011), while we might expect a linear exposure-response related to air pollution exposures. Finally, our measure of total active well production volume from both oil and gas did not distinguish whether the same production volume emanated from several or a few wells, as we were not able to account for differences in emissions impacts across different actively producing wells.

We observed some modest inverse associations between inactive wells and birth outcomes, primarily in urban areas. Inactive wells can pose risks in several ways. To date, excessive fugitive methane emissions have been measured at abandoned (unplugged) well sites, with higher concentrations detected at sites with compromised wells (Boothroyd et al. 2016; Kang et al. 2016). Residual off-gassing of air contaminants such as BTEX could also occur, which has prompted the South Coast air district and DOGGR to begin to collect air toxics and VOCs emissions data (LACDPH 2018; SCAQMD 2019; California AB1328 | TrackBill p. 13). Of greater concern is contamination of potable water sources from subsurface leakage and migration of contaminants through abandoned or idle wells (Long et al. 2015a). In an assessment of groundwater contamination from OGD in Ohio and Texas over more than a decade, abandoned wells accounted for 22% (Ohio) and 14% (Texas) of contamination incidents (Ground Water Protection Council 2011). In CA, idle wells may be repurposed for wastewater disposal or later revitalized with new technologies (Walker 2011). Wells operating with old infrastructure pose greater risks of leakages through the well casing and cement barriers (Ingraffea et al. 2014). Hydraulic fracturing could also increase the risk of surface or groundwater contamination via abandoned wells due to hydrological pressure changes; in one rare incident an abandoned well in Pennsylvania produced a 30-foot geyser of brine and gas for more than a week after a nearby gas well underwent hydraulic fracturing (US EPA 2016). We may not have observed any consistent or significant associations between exposure to inactive wells and adverse birth outcomes because we were not able to capture these nuanced exposure pathways with well count alone, leading to potential exposure misclassification.

Other limitations include our inability to adjust for several individual-level factors. Due to lack of data linkage, we could not control for the correlation between siblings (though we do include parity in all models) or maternal mobility during pregnancy. Birth records did not include a linking variable for siblings and only documented the residential address at time of birth. Previous studies on impacts of residential mobility during pregnancy suggest that ignoring residential mobility may lead to modest bias in associations towards the null or result in non-differential exposure misclassification (Chen et al. 2010; Hodgson et al. 2015; Lupo et al. 2010; Pennington et al. 2017). However, exposure estimates based on addresses captured at birth versus conception have been highly correlated (Chen et al. 2010; Lupo et al. 2010; Pennington et al. 2017). Across studies, ≤30% of mothers moved during pregnancy and moving distances were relatively short and within the same county (Bell and Belanger 2012; Chen et al. 2010; Hodgson et al. 2015; Lupo et al. 2010; Miller et al. 2010; Pennington et al. 2017). The extent of misclassification error depends on the spatial variability in the exposure (Hodgson et al. 2015). Additionally, exposure misclassification may be less prominent in the third trimester. Across environmental epidemiological studies that evaluated the impact of residential mobility on effect estimates by trimester, the highest rates of mobility occurred in the second trimester (Bell et al. 2018; Bell and Belanger 2012). Lowest residential mobility was observed in the first trimester among three studies and in the third trimester among two studies (Bell et al. 2018; Bell and Belanger 2012). Exposure misclassification due to mobility in the third trimester is less likely to be an issue, due to its proximity to the time of delivery, when the maternal residential address is collected and listed on the birth certificate. In addition to residential mobility, maternal occupational mobility should also be considered. One study that evaluated the impact of occupational mobility on air pollution exposure misclassification among Parisian women in the two first trimesters, found that mode of transport increased NO_2 exposure in the first trimester (Blanchard et al. 2018). Our study results yielded similar effect estimates across trimesters, suggesting that any bias resulting from maternal residential and occupational mobility is likely non-differential across trimesters.

In summary, this study expands the current literature on the health implications of OGD. We observed that prenatal exposure to active oil/gas production from both conventional and unconventional wells in CA was associated with adverse birth outcomes, and these associations varied by rural and urban areas. We observed the strongest associations with exposure to high production volume in rural areas. Future studies should consider inactive wells and conduct exposure assessments that collect environmental samples of OGD-related hazards. Such data would greatly improve exposure assignment and advance our understanding of underlying exposure sources and pathways. Additional evaluations of the relationship between oil/gas operator size, pollutant emissions, frequency and type of violations, and health outcomes would also elucidate which types of wells may be of greatest concern. Such data can inform regulatory decisions in terms of prioritizing inspection and pollution monitoring as well as emissions reduction requirements and community exposure reduction strategies.

1.6 Tables

Table 1.1. Neonate, maternal and area-level characteristics of births by oil and gas well production volume category, California 2006-2015. Pre-pregnancy BMI and smoking during pregnancy were available for 2007-2015 births (2006 births excluded from the missing category).

		Production volume			
Variable	**N (%)**	**No BOE (n=2,866,735)**	**1-100 BOE/day (n=70,615)**	**GT 100 BOE/day (n=50,079)**	**p-value†**
Neonate characteristics					
Mean birth weight (g) (SD)	2,987,429 (100)	3,327 ±528	3,318 ±527	3,316 ±527	<0.0001
Mean gestational age (weeks) (SD)	2,987,429 (100)	39 ±2	39 ±2	39 ±2	<0.0001
Sex					
Female	1,456,548 (49)	49	48	49	0.2879
Male	1,530,866 (51)	51	52	51	
Missing[a]	15 (<1)	100	0	0	
Birth month					
January	244,433 (8)	8	8	8	0.3261
February	224,691 (8)	8	8	8	
March	245,683 (8)	8	8	8	
April	233,297 (8)	8	8	8	
May	242,652 (8)	8	8	8	
June	241,962 (8)	8	8	8	
July	260,028 (9)	9	9	9	
August	269,714 (9)	9	9	9	
September	266,586 (9)	9	9	9	
October	261,399 (9)	9	9	9	
November	245,566 (8)	8	8	8	
December	251,418 (8)	8	8	8	
Birth year					
2006	320,330 (11)	11	10	12	<0.0001
2007	320,698 (11)	11	11	12	
2008	312,732 (10)	10	10	11	
2009	300,201 (10)	10	10	10	
2010	290,469 (10)	10	10	10	
2011	288,006 (10)	9	10	9	
2012	288,855 (9)	10	10	9	
2013	287,425 (10)	10	10	9	
2014	293,637 (10)	10	10	9	
2015	285,076 (10)	9	9	9	
Maternal Characteristics (%)					
Education					
< High school	764,090 (26)	26	31	21	<0.0001
High school diploma/GED	764,206 (26)	26	23	21	

Some college	724,574 (25)	25	22	23	
College+	665,993 (23)	23	24	35	
Missing[a]	68,566 (2)	95	3	2	
Age at delivery					
< 20	252,857 (8)	9	9	6	<0.0001
20-24	651,062 (22)	22	21	18	
25-29	809,072 (27)	27	27	25	
30-34	754,714 (25)	25	26	29	
35+	519,700 (17)	17	17	22	
Missing[a]	24 (<1)	92	8	0	
Race/ethnicity					
Asian/Pacific Islander	356,603 (12)	12	11	13	<0.0001
Black	154,047 (5)	5	6	9	
Hispanic	1,673,517 (56)	56	59	47	
Other	84,384 (3)	3	2	4	
White	718,878 (24)	24	22	27	
Kotelchuck index					
Inadequate	351,729 (12)	12	13	12	<0.0001
Intermediate	349,946 (12)	12	12	9	
Adequate+	905,545 (30)	30	29	34	
Adequate	1,380,209 (46)	46	46	45	
Parity					<0.0001
Nulliparous	1,154,875 (39)	39	40	44	
Multiparous	1,831,556 (61)	61	60	56	
Missing[a]	998 (<1)	93	4	3	
Mean pre-pregnancy BMI[b] (SD)	2,472,066 (93)	26 ±6	26 ±6	25 ±6	<0.0001
Missing[a]	195,033 (7)	94	4	2	
Smoking during pregnancy[b]					<0.0001
Smoked	49,461 (2)	2	1	1	
Did not smoke	257,7903 (97)	98	99	99	
Missing[a]	39,735 (1)	92	5	3	
TRI facility: 1+ within 1 km	48,189 (2)	2	4	3	<0.0001
Area-level characteristics (%)					
Mean NO_2 (ppb) (SD)	2,987,408 (99)	16 ±7	18 ±7	19 ±5	<0.0001
Missing[a]	21 (<1)	95	0	5	
Urban	2,651,066 (89)	89	87	97	
Air Basin					
Sacramento Valley	296,668 (10)	10	1	0.5	<0.0001
San Joaquin Valley	563,276 (19)	19	21	4	
South Central Coast	178,647 (6)	6	6	1	
South Coast	1,948,838 (65)	65	72	94	
ICE					
Quartile 1 - poverty	731,431 (25)	25	31	27	<0.0001
Quartile 2	731,403 (25)	25	23	19	
Quartile 3	730,283 (25)	25	19	23	

Quartile 4 - wealth	724,972 (25)	25	27	31	
Missing[a]	217 (<1)	76	9	15	
Oil/gas wells					
Mean inactive well count (SD)	2,987,429 (100)	0	89 ±111	160 ±191	<0.0001
Mean active well count	2,987,429 (100)	0	4 ±4	32 ±27	<0.0001
Mean production volume (BOE)/day (SD)	2,987,429 (100)	0	26 ±26	599 ±711	<0.0001

Note: BOE, barrels of oil equivalent; ICE: Index of Concentration at the Extremes.

†ANOVA or chi-square test

[a]Distribution of missingness across categories of production volume rather than percent missing in each production volume category.

[b]No covariate data available for 2006 (not included as missing), N=2,667,099 births between 2007 and 2015

1.7 Figures

Figure 1.1. Flow diagram of study population development and exclusion criteria applied.

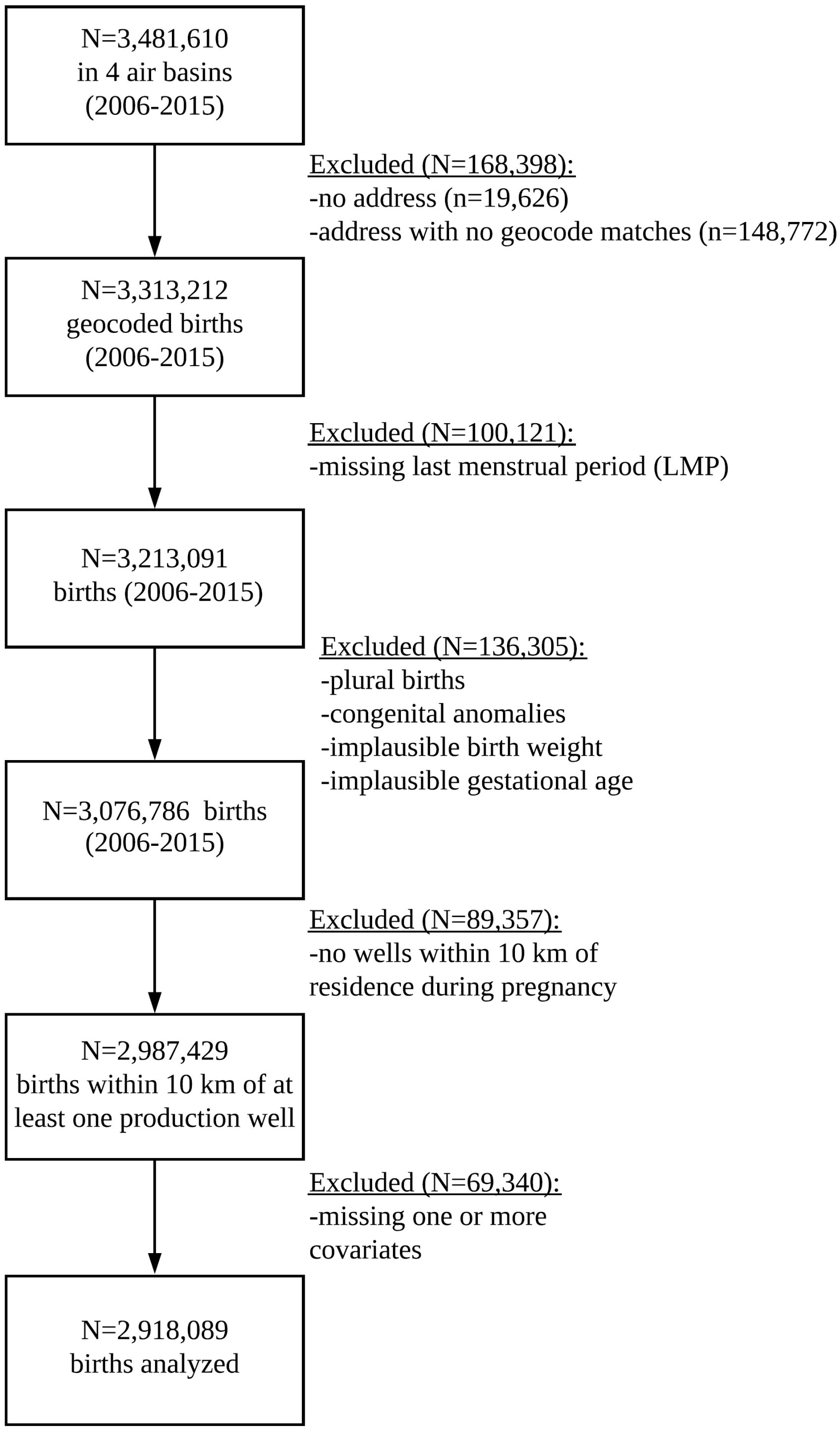
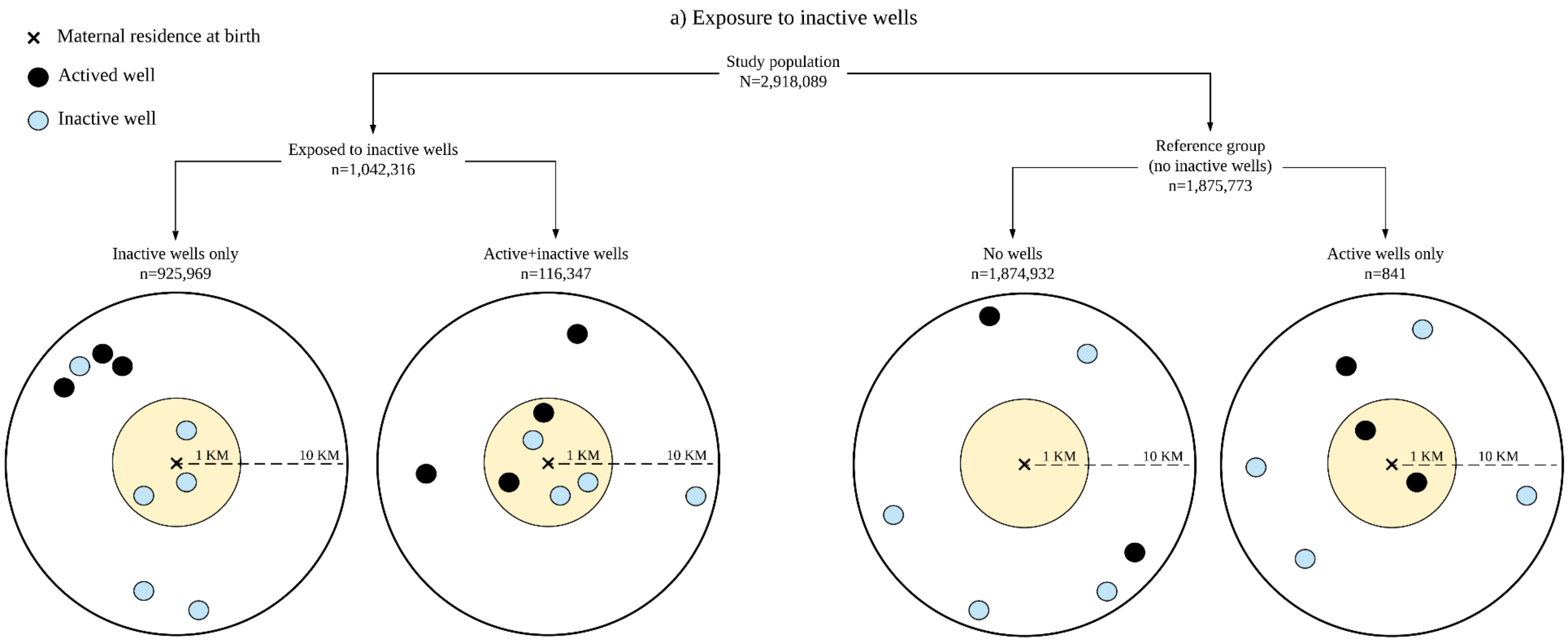


Figure 1.2. Schematic of definition of exposure and reference groups for inactive well count (A) and active well production volume (B). For each exposure metric, exposure was based on the presence of inactive or active wells within the 1 km buffer. Observations without the specific well type for each metric were assigned into the reference category.



b) Exposure to production volume from active wells

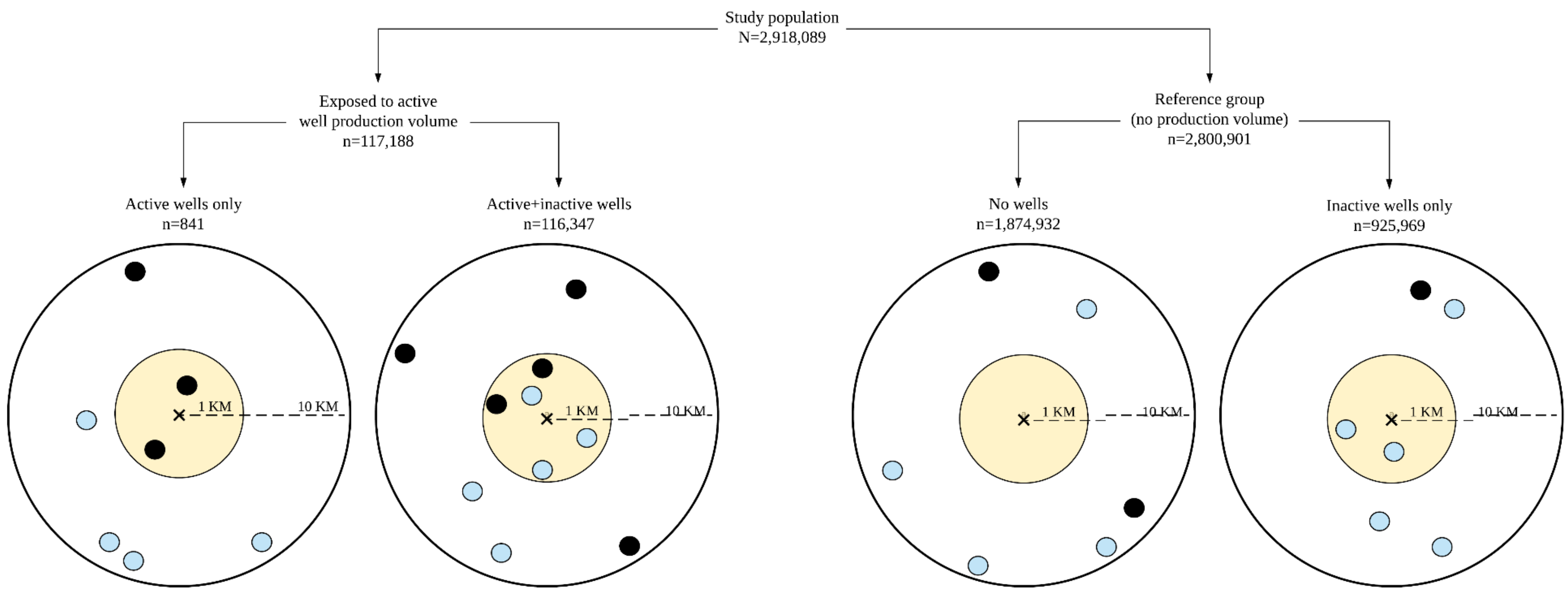
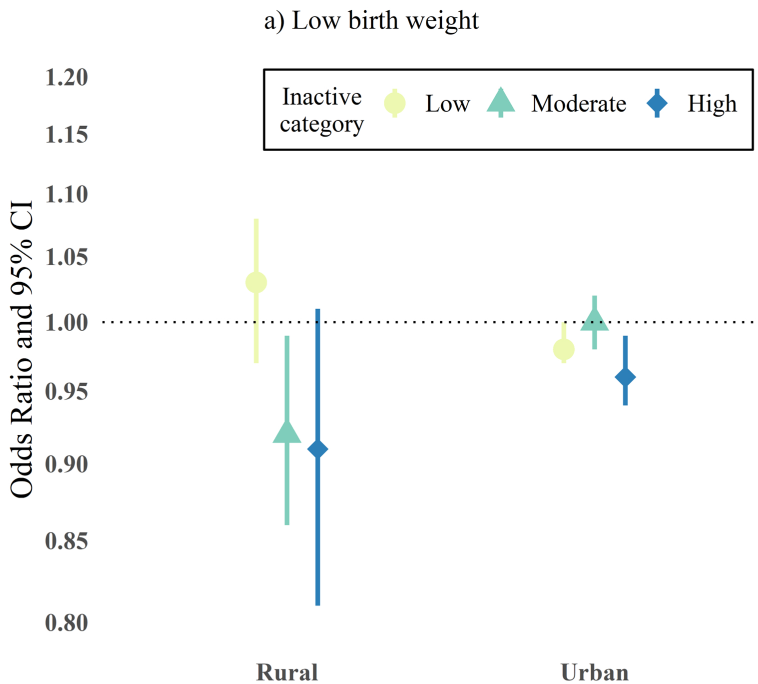
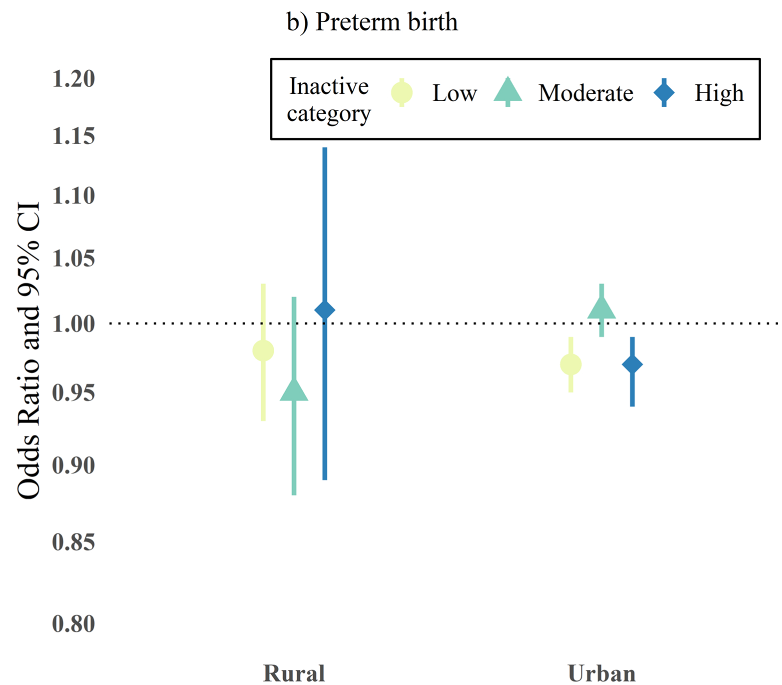


Figure 1.3. Plots of rural vs. urban odds ratios or mean difference in birth weight (grams) and 95% confidence interval for associations between exposure to low, moderate and high counts of inactive wells across the entire pregnancy and low birth weight (A), preterm birth (B), small for gestational age (C), and continuous term birth weight (D). Logistic regression models adjust for inactive well count, child's sex, birth month and birth year, and maternal education, age, race/ethnicity, Kotelchuck prenatal care index, parity, air basin, NO2 and ICE for income. In addition to the covariates adjusted for in the logistic regression models, the linear regression models also adjusted for gestational age. Numerical values plotted here can be found along with estimates for the three trimesters and p-values for statistical tests for effect modification in Supplemental Tables 1.1-1.2.



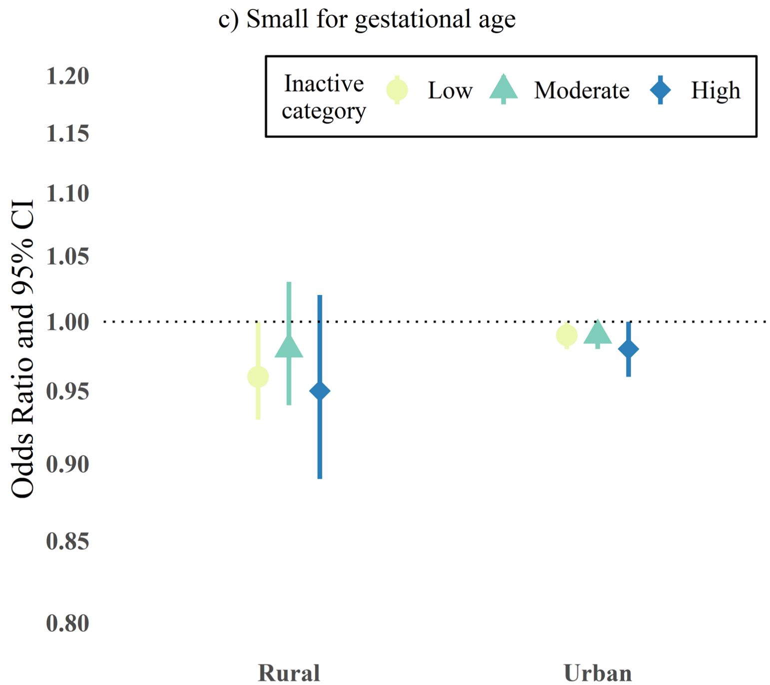
Number of births by inactive count category

	Rural	Urban
Low	36,357	380,234
Moderate	26,387	319,349
High	21,573	258,416



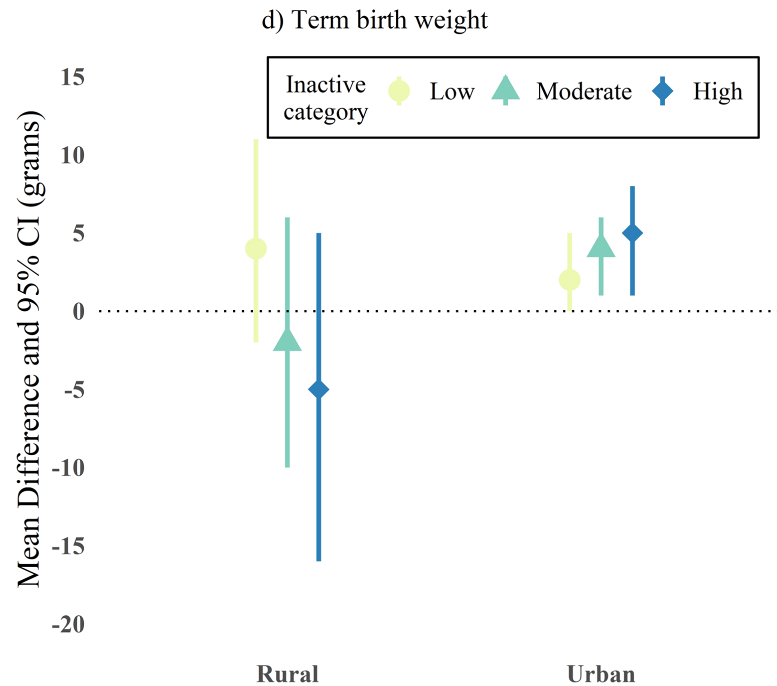
Number of births by inactive count category

	Rural	Urban
Low	36,357	380,234
Moderate	26,387	319,349
High	21,573	258,416



Number of births by inactive count category

	Rural	Urban
Low	36,357	380,234
Moderate	26,387	319,349
High	21,573	258,416

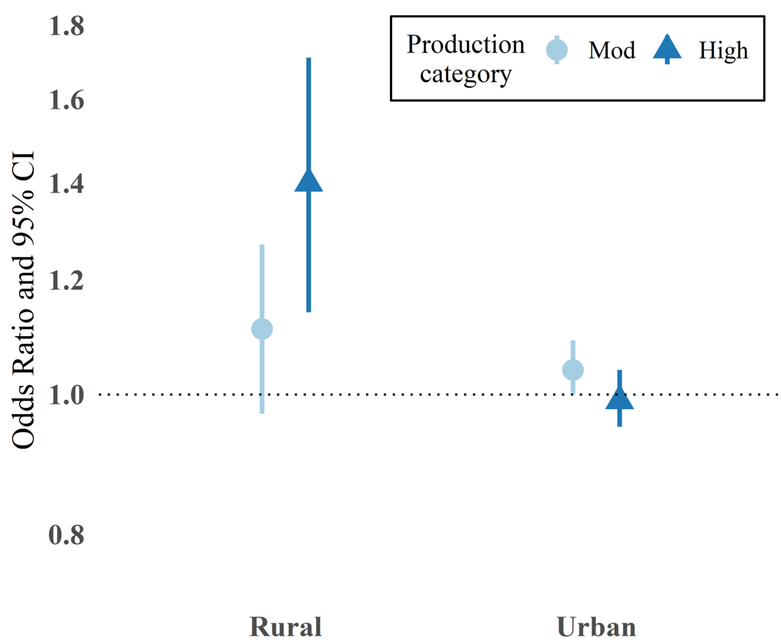


Number of births by inactive count category

	Rural	Urban
Low	34,056	354,859
Moderate	24,818	297,511
High	20,203	241,377

Figure 1.4. Plots of rural vs. urban odds ratios or mean difference in birth weight (grams) and 95% confidence interval for associations between exposure to moderate and high production volume across the entire pregnancy and low birth weight (A), preterm birth (B), small for gestational age (C), and continuous term birth weight (D). Logistic regression models adjust for inactive well count, child's sex, birth month and birth year, and maternal education, age, race/ethnicity, Kotelchuck prenatal care index, parity, air basin, NO2 and ICE for income. In addition to the covariates adjusted for in the logistic regression models, the linear regression models also adjusted for gestational age. Numerical values plotted here can be found along with estimates for the three trimesters and p-values for statistical tests for effect modification in Supplemental Tables 1.4-1.7.

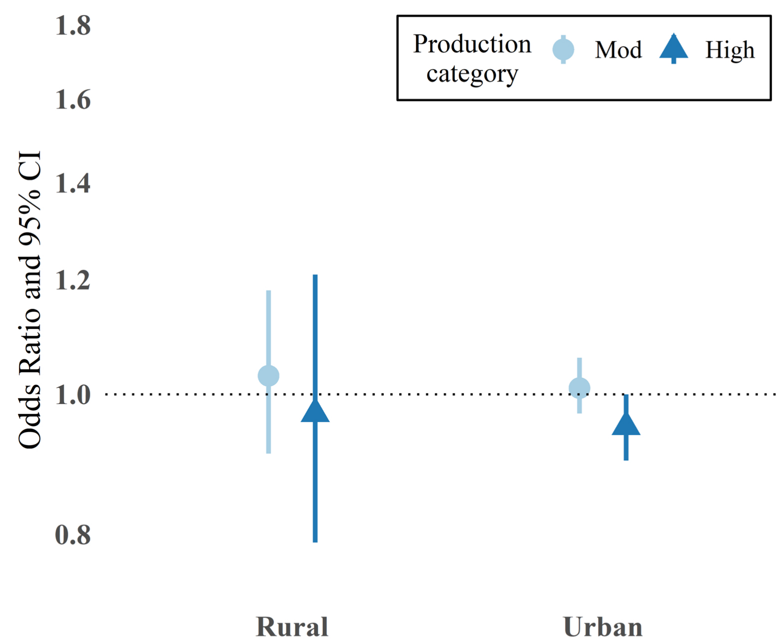
a) Low birth weight



Number of births by production category

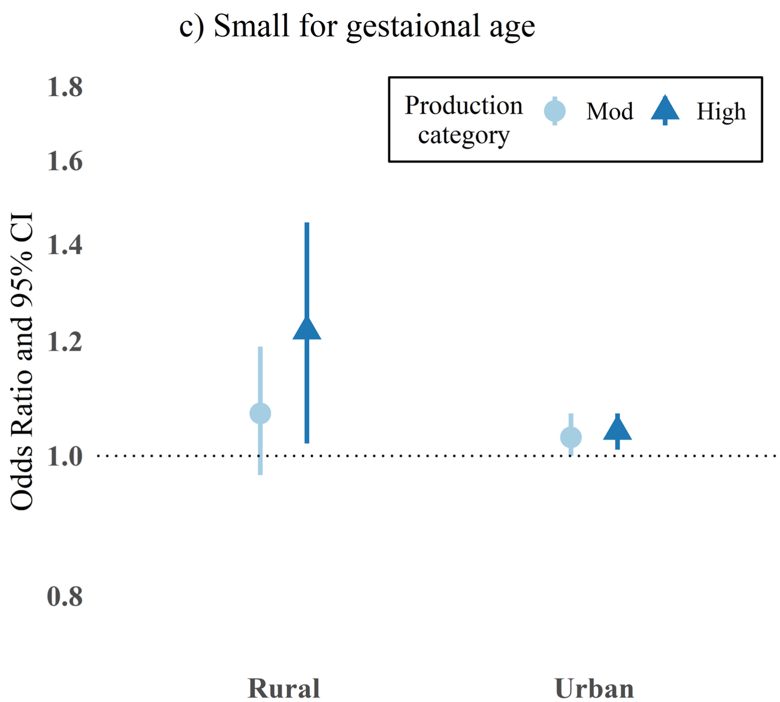
	Rural	Urban
Moderate	8,957	59,685
High	1,689	46,857

b) Preterm birth



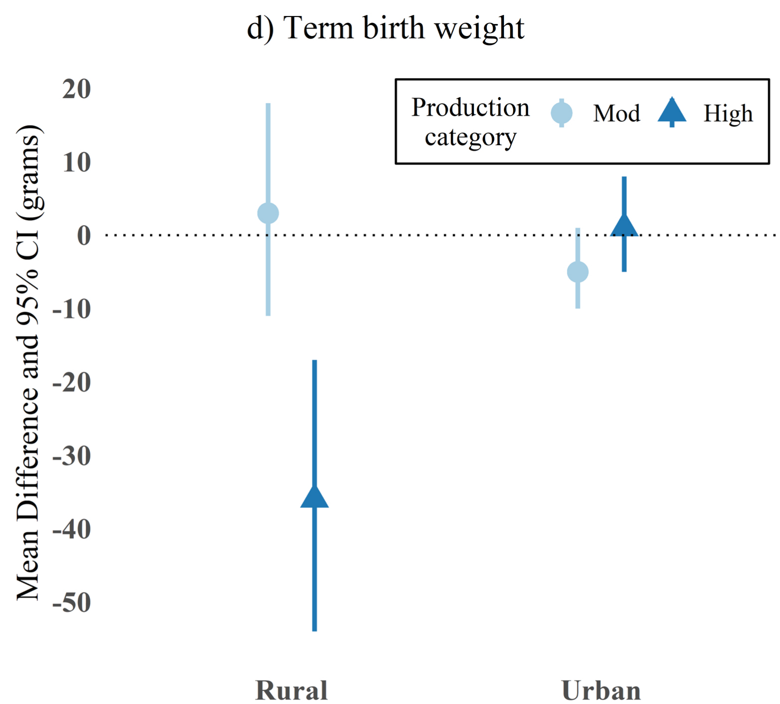
Number of births by production category

	Rural	Urban
Moderate	8,957	59,685
High	1,689	46,857



Number of births by production category

	Rural	Urban
Moderate	8,957	59,685
High	1,689	46,857

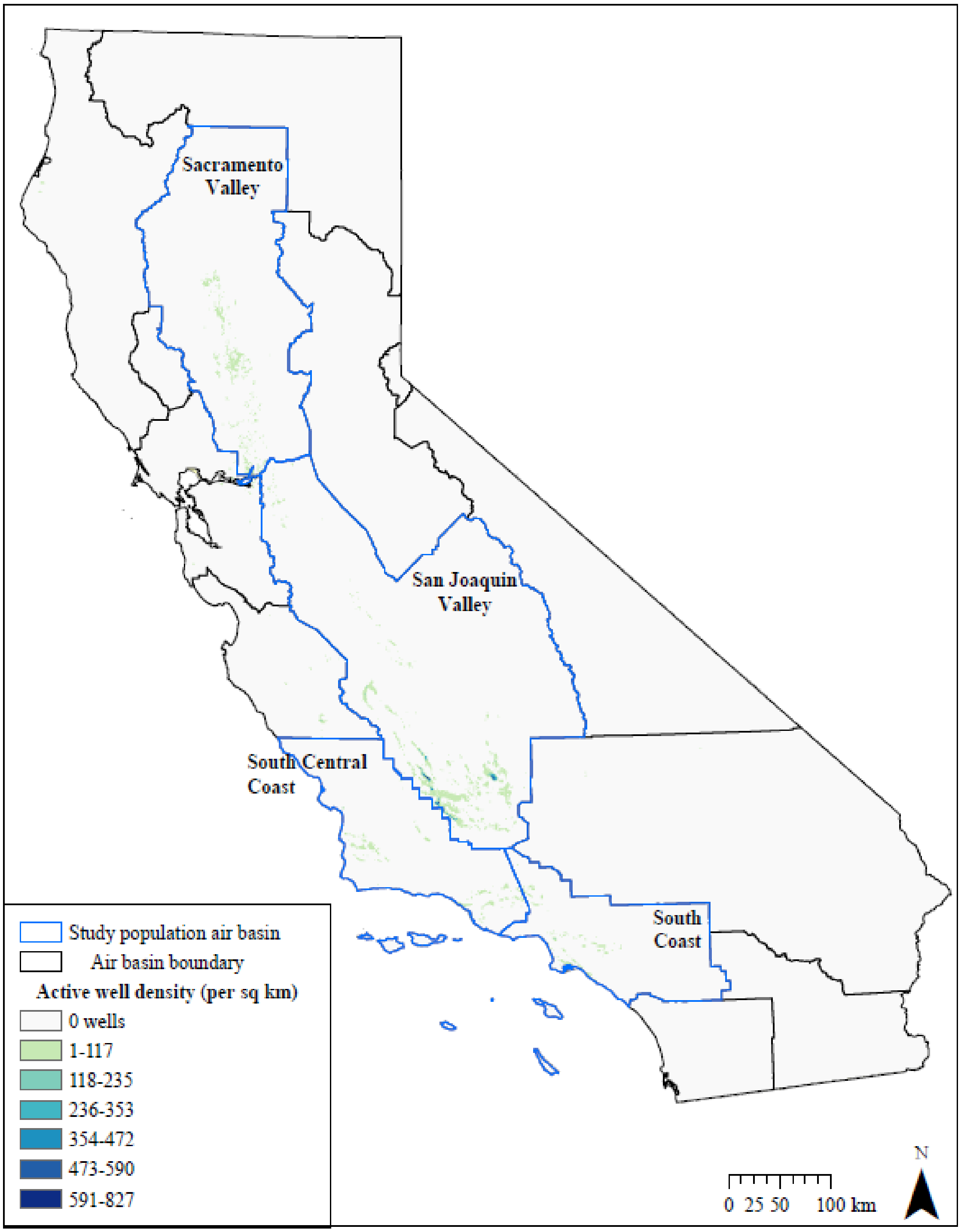


Number of births by production category

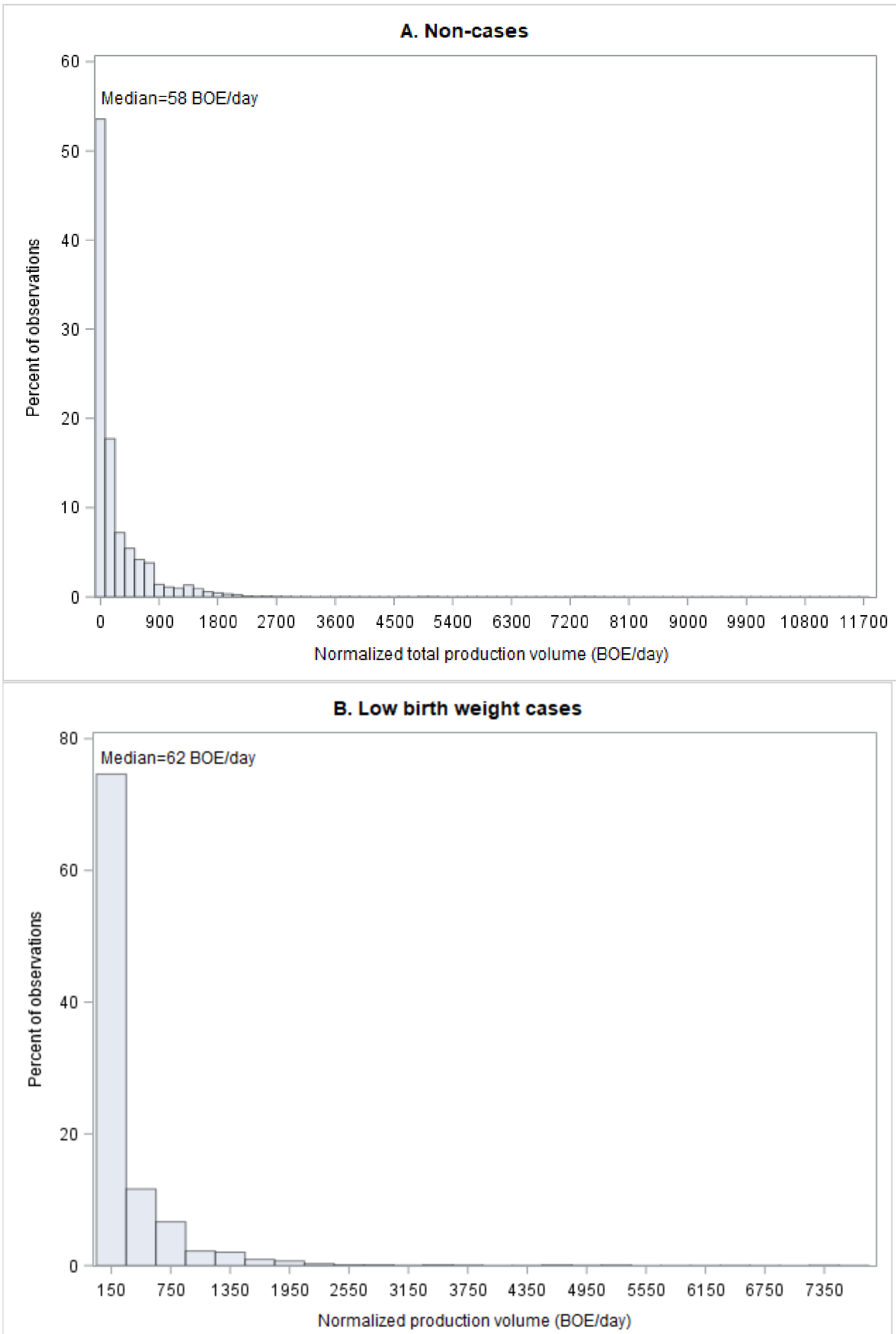
	Rural	Urban
Moderate	8,339	55,565
High	1,590	43,770

1.8 Supplemental Information Chapter 1

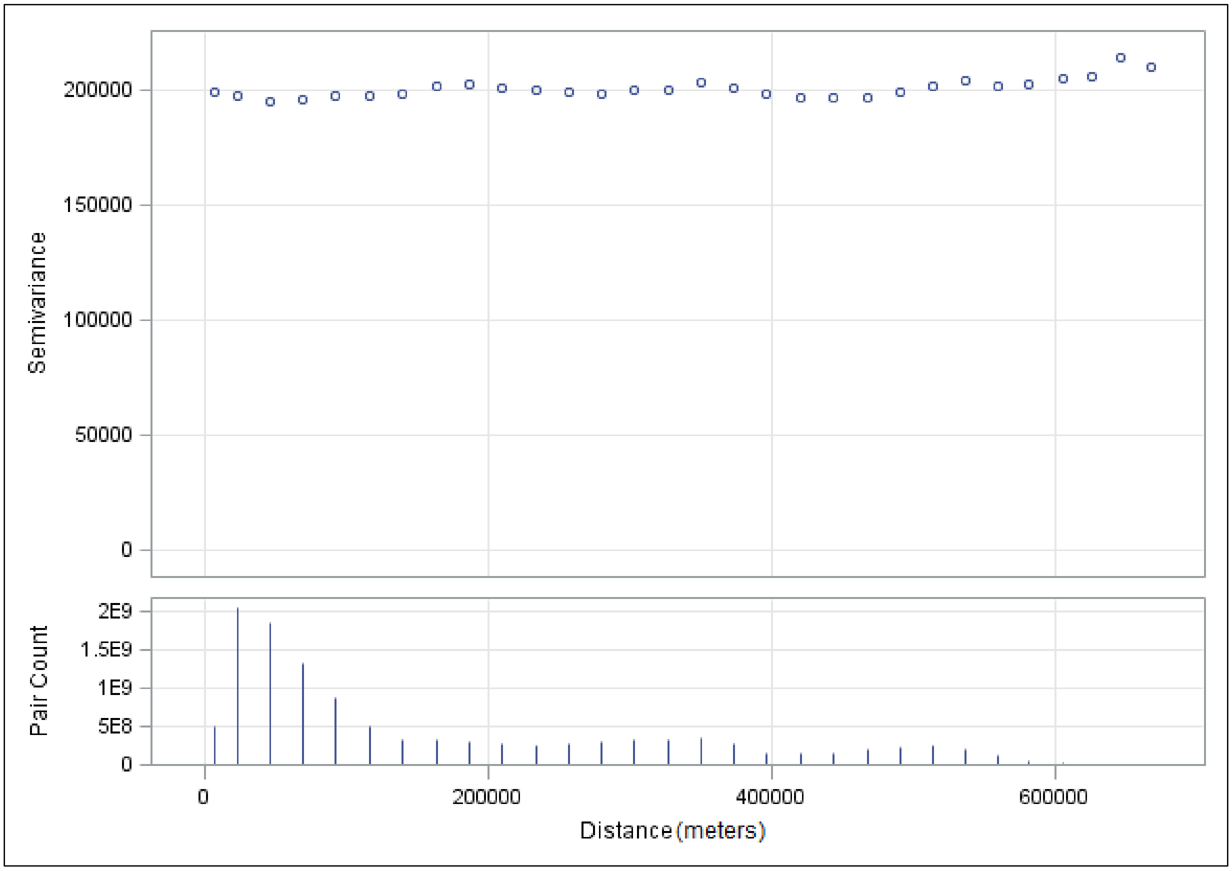
Supplemental Figure 1.1. Active well density by air basin across California (2005–2015). Map created in ArcGIS 10.6 (ESRI, Redlands, CA). The well density was calculated via the point density tool, which calculates density based on the number of neighboring wells within a 1 km x 1 km cell around each well by air basin.



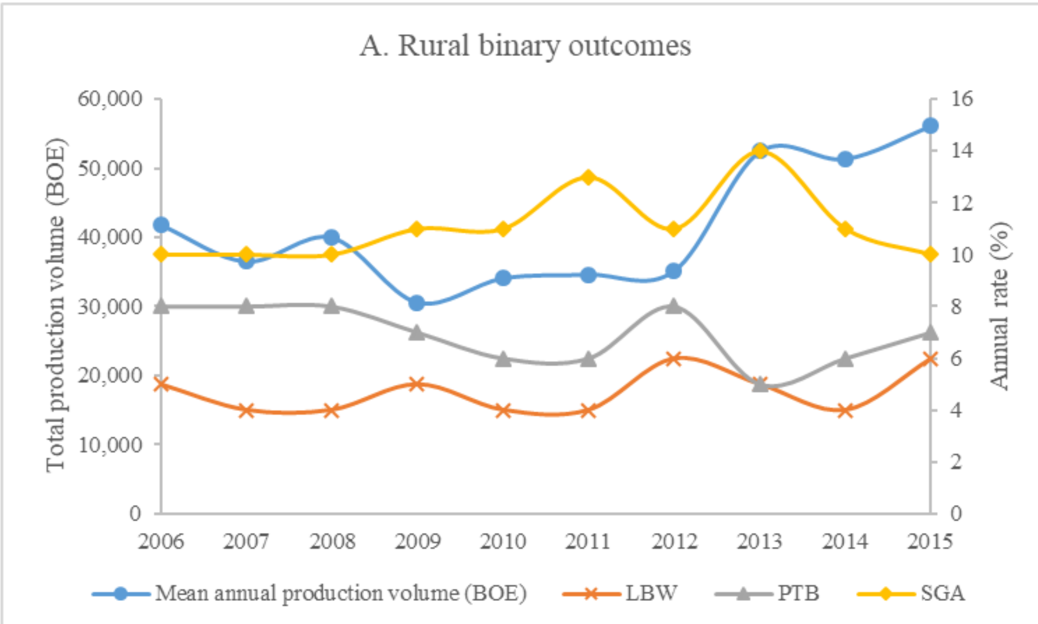
Supplemental Figure 1.2. Distribution of total production volume (BOE) per day by non-cases of LBW (A) and cases of LBW (B). The distribution was generated for birth outcome cases and non-cases in order to select a cut-off around the median for each category of exposure to active wells. The distribution of normalized BOE/day was similar across cases of LBW, PTB and SGA. One cut-off was selected for comparability and the value of the cut-off was selected to ensure sufficient overall sample size and number of cases in each exposure group. Figures generated in SAS 9.4 (SAS Institute Inc., Cary, NC).

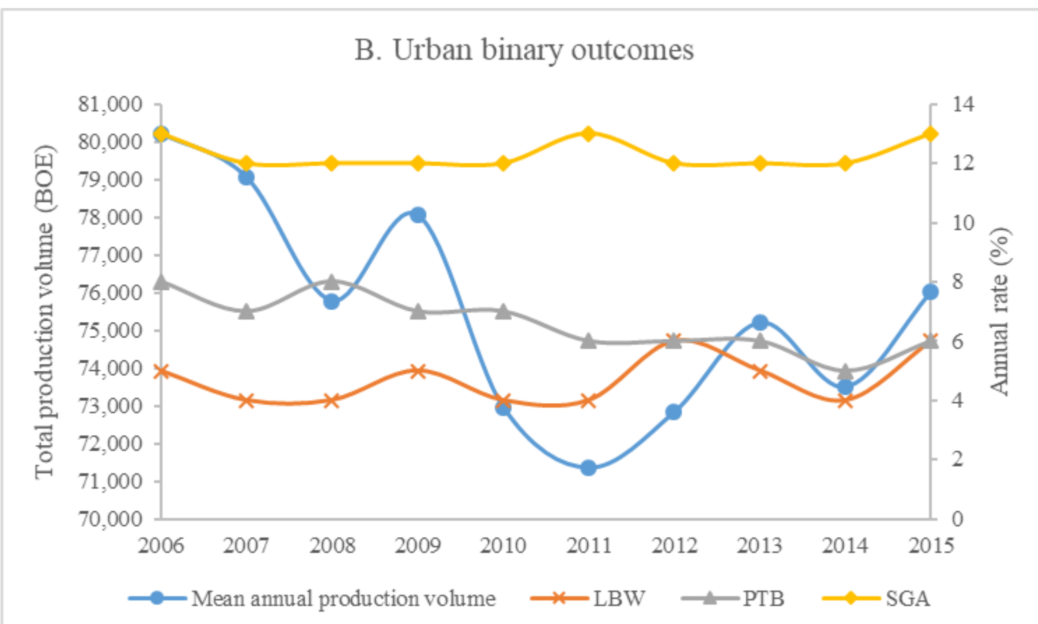


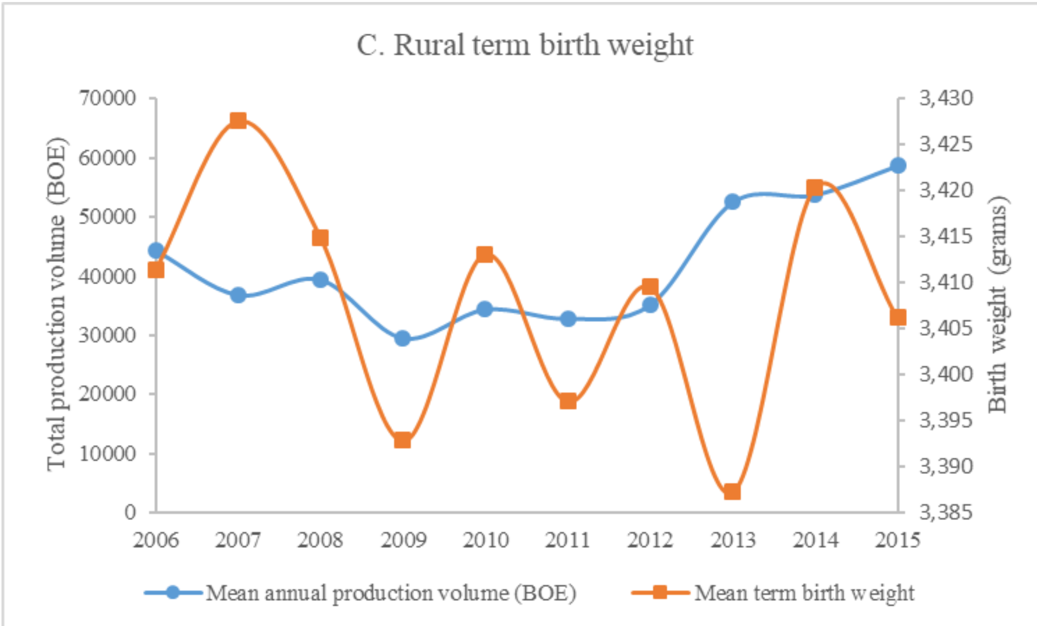
Supplemental Figure 1.3. Semi-variogram generated to assess residual spatial dependence. Distances between observations were grouped into 30 distance classes to generate corresponding averages of the semi-variances. The nearly straight line indicates that there is no spatial autocorrelation between the observations. Figure generated in SAS 9.4 (SAS Institute Inc., Cary, NC).

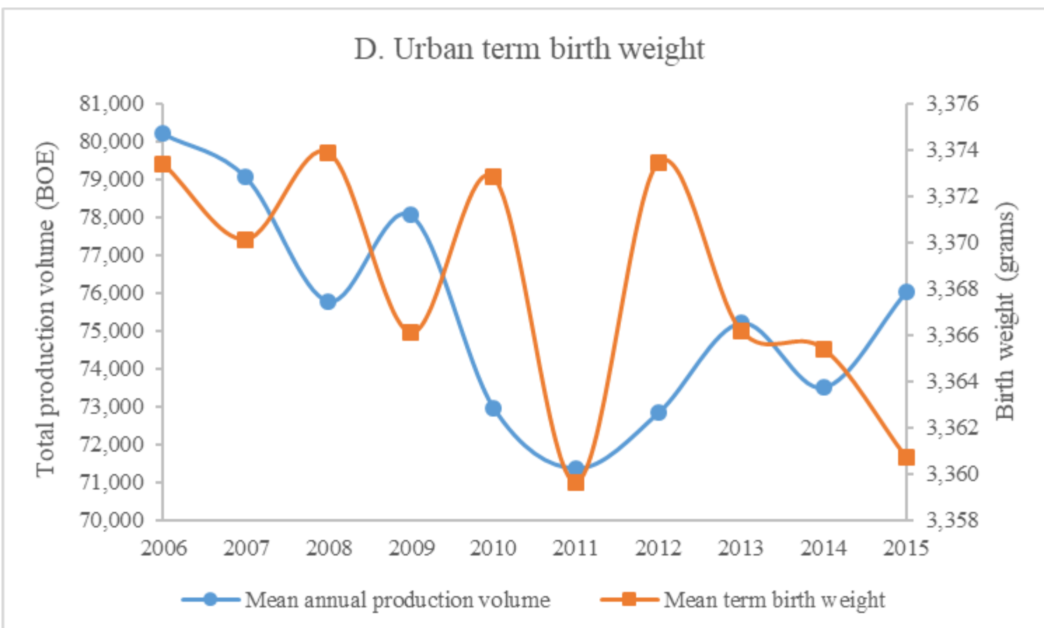


Supplemental Figure 1.4. Distribution of the average total production volume (BOE) and birth outcomes by year of birth. Total production volume reflects exposure that occurred during pregnancy, most of which occurred the prior year of birth. The annual rates of LBW, PTB and SGA are plotted for rural (A) and urban (B) areas. The birth weight plots include only term births and the mean term birth weight is plotted for rural (C) and urban (D) areas.






C. Rural term birth weight
Total production volume (BOE)
70000
60000
50000
40000
30000
20000
10000
0
3,430
3,425
3,420
3,415
3,410
3,405
3,400
3,395
3,390
3,385
Birth weight (grams)
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
Mean annual production volume (BOE)
Mean term birth weight


D. Urban term birth weight
Total production volume (BOE)
81,000
80,000
79,000
78,000
77,000
76,000
75,000
74,000
73,000
72,000
71,000
70,000
3,376
3,374
3,372
3,370
3,368
3,366
3,364
3,362
3,360
3,358
Birth weight (grams)
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
Mean annual production volume
Mean term birth weight

Supplemental Table 1.1. Adjusted odds ratio for binary birth outcomes associated with exposure to inactive wells by rural/urban status. Data for Figure 1.3A-C.

Inactive count categories	No BOE (ref)		1 well			2-5 wells			6+ wells		
	n	Cases (%)	*n*	Cases (%)	aOR (95% CI)	*n*	Cases (%)	aOR (95% CI)	n	Cases (%)	aOR (95% CI)
Low birth weight: Rural[a]											
Entire pregnancy	244,817	11,225 (5)	36,357	1,684 (5)	1.03 (0.97, 1.08)	26,387	1,116 (4)	0.92 (0.86, 0.99)	21,573	920 (4)	0.91 (0.81, 1.01)
Trimester 1	244,817	11,225 (5)	36,347	1,684 (5)	1.03 (0.97, 1.08)	26,371	1,116 (4)	0.92 (0.86, 0.99)	21,599	920 (4)	0.90 (0.81, 1.01)
Trimester 2	244,816	11,225 (5)	36,349	1,684 (5)	1.03 (0.97, 1.08)	26,380	1,115 (4)	0.92 (0.86, 0.99)	21,589	921 (4)	0.91 (0.82, 1.01)
Trimester 3	244,045	10,609 (4)	36,231	1,589 (4)	1.02 (0.97, 1.08)	26,307	1,058 (4)	0.92 (0.85, 0.99)	21,540	864 (4)	0.91 (0.82, 1.02)
Low birth weight: Urban[a]											
Entire pregnancy	1,630,956	84,068 (5)	380,234	19,391 (5)	0.98 (0.97, 1.00)	319,349	16,515 (5)	1.00 (0.98, 1.02)	258,416	13,181 (5)	0.96 (0.94, 0.99)
Trimester 1	1,630,889	84,063 (5)	380,265	19,391 (5)	0.98 (0.97, 1.00)	319,300	16,516 (5)	1.00 (0.98, 1.02)	258,501	13,185 (5)	0.96 (0.94, 0.99)
Trimester 2	1,630,924	84,067 (5)	380,233	19,391 (5)	0.98 (0.97, 1.00)	319,317	16,512 (5)	1.00 (0.98, 1.02)	258,481	13,185 (5)	0.96 (0.94, 0.99)
Trimester 3	1,625,253	79,173 (5)	378,963	18,278 (5)	0.98 (0.97, 1.00)	318,249	15,582 (5)	1.00 (0.98, 1.02)	257,713	12,483 (5)	0.97 (0.94, 1.00)
Preterm birth: Rural[a]											
Entire pregnancy	244,817	16,322 (7)	36,357	2,301 (6)	0.98 (0.93, 1.03)	26,387	1,569 (6)	0.95 (0.88, 1.02)	21,573	1,370 (6)	1.01 (0.89, 1.14)
Trimester 1	244,817	16,322 (7)	36,347	2,301 (6)	0.98 (0.93, 1.03)	26,371	1,567 (6)	0.94 (0.88, 1.02)	21,599	1,372 (6)	1.01 (0.89, 1.14)
Trimester 2	244,816	16,322 (7)	36,349	2,301 (6)	0.98 (0.93, 1.03)	26,380	1,567 (6)	0.94 (0.88, 1.02)	21,589	1,372 (6)	1.00 (0.88, 1.13)
Trimester 3	244,045	15,551 (6)	36,231	2,191 (6)	0.98 (0.93, 1.03)	26307	1,500 (6)	0.94 (0.88, 1.02)	21,540	1,309 (6)	1.02 (0.90, 1.15)
Preterm birth: Urban[a]											
Entire pregnancy	1,630,956	113,646 (7)	380,234	25,375 (7)	0.97 (0.95, 0.99)	319,349	21,838 (7)	1.01 (0.99, 1.03)	258,416	17,039 (7)	0.97 (0.94, 0.99)
Trimester 1	1,630,889	113,635 (7)	380,265	25,383 (7)	0.97 (0.96, 0.99)	319,300	21,835 (7)	1.01 (0.99, 1.03)	258,501	17,045 (7)	0.97 (0.94, 0.99)
Trimester 2	1,630,924	113,644 (7)	380,233	25,376 (7)	0.97 (0.95, 0.99)	319,317	21,835 (7)	1.01 (0.99, 1.03)	258,481	17,043 (7)	0.97 (0.94, 0.99)
Trimester 3	1,625,253	107,975 (7)	378,963	24,117 (6)	0.97 (0.95, 0.99)	318,249	20,787 (7)	1.01 (0.99, 1.03)	257,713	16,242 (6)	0.97 (0.95,1.00)
Small for gestational age: Rural[a]											
Entire pregnancy	244,817	25,536 (10)	36,357	3,702 (10)	0.96 (0.93, 1.00)	26,387	2,726 (10)	0.98 (0.94, 1.03)	21,573	2,247 (10)	0.95 (0.89, 1.02)
Trimester 1	244,817	25,536 (10)	36,347	3,701 (10)	0.96 (0.93, 1.00)	26,371	2,724 (10)	0.98 (0.94, 1.03)	21,599	2,250 (10)	0.96 (0.89, 1.03)
Trimester 2	244,816	25,536 (10)	36,349	3,702 (10)	0.96 (0.93, 1.00)	26,380	2,724 (10)	0.98 (0.93, 1.03)	21,589	2,249 (10)	0.95 (0.89, 1.02)
Trimester 3	244,045	25,501 (10)	36,231	3,699 (10)	0.96 (0.93, 1.00)	26307	2,722 (10)	0.98 (0.94, 1.03)	21,540	2,247 (10)	0.95 (0.89, 1.02)
Small for gestational age: Urban[a]											
Entire pregnancy	1,630,956	189,858 (12)	380,234	45,158 (12)	0.99 (0.98, 1.00)	319,349	37,830 (12)	0.99 (0.98, 1.00)	258,416	30,886 (12)	0.98 (0.96, 1.00)
Trimester 1	1,630,889	189,846 (12)	380,265	45,165 (12)	0.99 (0.98, 1.00)	319,300	37,825 (12)	0.99 (0.98, 1.00)	258,501	30,896 (12)	0.98 (0.96, 1.00)
Trimester 2	1,630,924	189,856 (12)	380,233	45,157 (12)	0.99 (0.98, 1.00)	319,317	37,822 (12)	0.99 (0.98, 1.00)	258,481	30,897 (12)	0.98 (0.96, 1.00)
Trimester 3	1,625,253	189,539 (12)	378,963	45,094 (12)	0.99 (0.98, 1.00)	318,249	37,773 (12)	0.99 (0.98, 1.00)	257,713	30,849 (12)	0.98 (0.96, 1.00)

Note: aOR, adjusted odds ratio, CI, confidence interval.

[a]Logistic regression models adjusted for production volume; child's sex, birth month and birth year; maternal education, age, race/ethnicity, Kotelchuck prenatal care index, parity; air basin, NO_2 concentration, and ICE for income.

Supplemental Table 1.2. Adjusted odds ratio for term birth weight (grams) associated with exposure to inactive wells by rural/urban status. Data for Figure 1.3D.

Inactive count categories	0 wells (ref)	1 well		2-5 wells		6+ wells	
	n	*n*	aDiff (95% CI)	*n*	aDiff (95% CI)	n	aDiff (95% CI)
Rural[a]							
Entire pregnancy	228,495	34,056	4 (-2, 11)	24,818	-2 (-10, 6)	20,203	-5 (-16, 5)
Trimester 1	228,495	34,046	4 (-2, 11)	24,804	-2 (-10, 6)	20,227	-5 (-16, 5)
Trimester 2	228,494	34,048	4 (-2, 11)	24,813	-2 (-10, 6)	20,217	-5 (-16, 5)
Trimester 3	228,494	34,040	4 (-2, 11)	24,807	-2 (-10, 5)	20,231	-6 (-17, 5)
Urban[a]							
Entire pregnancy	1,517,310	354,859	2 (0, 5)	297,511	4 (1, 6)	241,377	5 (1, 8)
Trimester 1	1,517,254	354,882	2 (0, 5)	297,465	4 (1, 6)	241,456	5 (1, 8)
Trimester 2	1,517,280	354,857	2 (0, 5)	297,482	4 (1, 6)	241,438	5 (1, 8)
Trimester 3	1,517,278	354,846	2 (0, 5)	297,462	4 (1, 6)	241,471	5 (1, 8)

Note: aDiff, adjusted mean difference (grams); CI, confidence interval.

[a]Linear regression models adjusted for production volume; child's gestational age, sex, birth month and birth year; maternal education, age, race/ethnicity, Kotelchuck prenatal care index, parity; air basin, NO_2 concentration, and ICE for income.

Supplemental Table 1.3. Unstratified adjusted odds ratios and mean difference (grams) for adverse birth outcomes associated with exposure to oil and gas production.

Prod volume categories	No BOE (ref)		1-100 BOE/day			GT 100 BOE/day		
	n	Cases (%)	*n*	Cases (%)	EE (95% CI)	n	Cases (%)	EE (95% CI)
Low birth weight[a]								
Entire pregnancy	2,800,901	141,984 (5)	68,642	3,561 (5)	1.05 (1.01, 1.09)	48,546	2,555 (5)	1.00 (0.96, 1.06)
Trimester 1	2,801,853	142,033 (5)	67,776	3,513 (5)	1.05 (1.00, 1.09)	48,460	2,554 (5)	1.01 (0.96, 1.06)
Trimester 2	2,801,831	142,027 (5)	63,706	3,317 (5)	1.05 (1.01, 1.10)	52,552	2,756 (5)	1.01 (0.96, 1.05)
Trimester 3	2,793,270	133,973 (5)	72,835	3,657 (6)	1.06 (1.02, 1.11)	42,196	2,006 (4)	0.94 (0.89, 1.00)
Preterm birth[a]								
Entire pregnancy	2,800,901	191,536 (7)	68,642	4,738 (7)	1.01 (0.97, 1.06)	48,546	3,186 (7)	0.95 (0.91, 1.00)
Trimester 1	2,801,853	191,592 (7)	67,776	4,692 (7)	1.02 (0.98, 1.06)	48,460	3,176 (7)	0.95 (0.91, 1.00)
Trimester 2	2,801,831	191,578 (7)	63,706	4,450 (7)	1.03 (0.99, 1.07)	52,552	3,432 (7)	0.95 (0.90, 1.00)
Trimester 3	2,793,270	182,284 (7)	72,835	5,011 (7)	1.06 (1.02, 1.10)	42,196	2,377 (6)	0.83 (0.78, 0.88)
Small for gestational age[a]								
Entire pregnancy	2,800,901	323,688 (12)	68,642	8,305 (12)	1.04 (1.00, 1.07)	48,546	5,950 (12)	1.05 (1.01, 1.08)
Trimester 1	2,801,853	323,824 (12)	67,776	8,183 (12)	1.03 (1.00, 1.07)	48,460	5,936 (12)	1.05 (1.01, 1.08)
Trimester 2	2,801,831	323,806 (12)	63,706	7,723 (12)	1.04 (1.00, 1.07)	52,552	6,414 (12)	1.04 (1.01, 1.07)
Trimester 3	2,793,270	323,405 (12)	72,835	8,806 (12)	1.04 (1.00, 1.07)	42,196	5,213 (12)	1.05 (1.02, 1.09)
Term birth weight[b]								
Entire pregnancy	2,609,365	--	63,904	--	-3 (-8, 2)	45,360	--	-5 (-10, 1)
Trimester 1	2,610,261	--	63,084	--	-3 (-8, 2)	45,284	--	-4 (-10, 2)
Trimester 2	2,610,253	--	59,256	--	-3 (-8, 3)	49,120	--	-4 (-10, 1)
Trimester 3	2,610,986	--	67,824	--	-3 (-8, 2)	39,819	--	-4 (-10, 3)

Note: EE, effect estimate; CI, confidence interval; BOE, barrel of oil equivalents of oil and gas; GT, greater than.

[a]Logistic regression models (odds ratio) adjusted for inactive well count; child's sex, birth month and birth year; maternal education, age, race/ethnicity, Kotelchuck prenatal care index, parity; urban indicator, air basin, NO_2 concentration, and ICE for income.

[b]Linear regression model (mean difference, grams) also adjusted for gestational age.

Supplemental Table 1.4. Adjusted odds ratios for low birth weight associated with oil and gas production volume by urban/rural status. Data for Figure 1.4A. Effect modification p-values were derived from two-sample z-tests using strata-specific estimates and variances; significance at α=0.05.

Prod volume categories	No BOE (ref)		1-100 BOE/day				GT 100 BOE/day			
	n	Cases (%)	*n*	Cases (%)	aOR (95% CI)	EM p-value	n	Cases (%)	aOR (95% CI)	EM p-value
Rural[a]										
Entire pregnancy	318,488	14,451 (5)	8,957	400 (4)	1.11 (0.97, 1.27)	0.81	1,689	94 (6)	1.40 (1.14, 1.71)	0.01
Trimester 1	318,629	14,457 (5)	8,809	394 (4)	1.12 (0.98, 1.28)	0.67	1,696	94 (6)	1.39 (1.11, 1.75)	0.002
Trimester 2	318,675	14,461 (5)	8,258	367 (4)	1.10 (0.96, 1.26)	1.00	2,201	117 (5)	1.35 (1.13, 1.61)	0.002
Trimester 3	317,913	13,684 (4)	8,790	359 (4)	1.07 (0.93, 1.23)	1.00	1,420	77 (5)	1.38 (1.11, 1.72)	0.01
Urban[a]										
Entire pregnancy	2,482,413	127,533 (5)	59,685	3,161 (5)	1.04 (1.00, 1.09)	--	46,857	2,461 (5)	0.99 (0.95, 1.04)	--
Trimester 1	2,483,224	127,576 (5)	58,967	3,119 (5)	1.04 (0.99, 1.09)	--	46,764	2,460 (5)	1.00 (0.95, 1.04)	--
Trimester 2	2,483,156	127,566 (5)	55,448	2,950 (5)	1.05 (1.00, 1.10)	--	50,351	2,639 (5)	0.99 (0.95, 1.04)	--
Trimester 3	2,475,357	120,289 (5)	64,045	3,298 (5)	1.06 (1.02, 1.11)	--	40,776	1,929 (5)	0.93 (0.88, 0.98)	--

Note: aOR, adjusted odds ratio; CI, confidence interval; BOE, barrel of oil equivalents of oil and gas; GT, greater than; EM, effect modification.

[a]Logistic regression models adjusted for inactive well count; child's sex, birth month and birth year; maternal education, age, race/ethnicity, Kotelchuck prenatal care index, parity; air basin, NO_2 concentration, and ICE for income.

Supplemental Table 1.5. Adjusted odds ratios for preterm birth associated with oil and gas production volume by urban/rural status. Data for Figure 1.4B. Effect modification p-values were derived from two-sample z-tests using strata-specific estimates and variances; significance at α=0.05.

Prod volume categories	No BOE (ref)		1-100 BOE/day				GT 100 BOE/day			
	n	Cases (%)	*n*	Cases (%)	aOR (95% CI)	EM p-value	n	Cases (%)	aOR (95% CI)	EM p-value
Rural[a]										
Entire pregnancy	318,488	20,845 (7)	8,957	618 (7)	1.03 (0.91, 1.18)	1.00	1,689	99 (6)	0.97 (0.78, 1.21)	1.00
Trimester 1	318,629	20,857 (7)	8,809	604 (7)	1.02 (0.90, 1.16)	1.00	1,696	101 (6)	1.00 (0.80, 1.24)	1.00
Trimester 2	318,675	20,850 (7)	8,258	582 (7)	1.06 (0.94, 1.21)	1.00	2,201	130 (6)	0.98 (0.82, 1.18)	1.00
Trimester 3	317,913	19,899 (6)	8,790	575 (7)	1.03 (0.90, 1.17)	1.00	1,420	77 (5)	0.92 (0.71, 1.19)	0.84
Urban[a]										
Entire pregnancy	2,482,413	170,691 (7)	59,685	4,120 (7)	1.01 (0.97, 1.06)	--	46,857	3,087 (7)	0.95 (0.90, 1.00)	--
Trimester 1	2,483,224	170,735 (7)	58,967	4,088 (7)	1.01 (0.97, 1.06)	--	46,764	3,075 (7)	0.95 (0.91, 1.00)	--
Trimester 2	2,483,156	170,728 (7)	55,448	3,868 (7)	1.02 (0.98, 1.07)	--	50,351	3,302 (7)	0.95 (0.90, 1.00)	--
Trimester 3	2,475,357	162,385 (7)	64,045	4,436 (7)	1.06 (1.02, 1.11)	--	40,776	2,300 (6)	0.82 (0.77, 0.88)	--

Note: aOR, adjusted odds ratio; CI, confidence interval; BOE, barrel of oil equivalents of oil and gas; GT, greater than; EM, effect modification.

[a]Logistic regression models adjusted for inactive well count; child's sex, birth month and birth year; maternal education, age, race/ethnicity, Kotelchuck prenatal care index, parity; air basin, NO_2 concentration, and ICE for income.

Supplemental Table 1.6. Adjusted odds ratios for small for gestational age associated with oil and gas production volume by urban/rural status. Data for Figure 1.4C. Effect modification p-values were derived from two-sample z-tests using strata-specific estimates and variances; significance at $\alpha=0.05$.

Prod volume categories	No BOE (ref)		1-100 BOE/day				GT 100 BOE/day			
	n	Cases (%)	*n*	Cases (%)	aOR (95% CI)	EM p-value	n	Cases (%)	aOR (95% CI)	EM p-value
Rural[a]										
Entire pregnancy	318,488	33,034 (10)	8,957	966 (11)	1.07 (0.97, 1.19)	0.99	1,689	211 (13)	1.22 (1.02, 1.45)	0.14
Trimester 1	318,629	33,056 (10)	8,809	937 (11)	1.05 (0.95, 1.16)	1.00	1,696	218 (13)	1.25 (1.04, 1.50)	0.07
Trimester 2	318,675	33,058 (10)	8,258	889 (11)	1.07 (0.96, 1.19)	1.00	2,201	264 (12)	1.17 (1.02, 1.35)	0.20
Trimester 3	317,913	33,038 (10)	8,790	948 (11)	1.08 (0.97, 1.19)	0.90	1,420	183 (13)	1.24 (1.02, 1.50)	0.14
Urban[a]										
Entire pregnancy	2,482,413	290,654 (12)	59,685	7,339 (12)	1.03 (1.00, 1.07)	--	46,857	5,739 (12)	1.04 (1.01, 1.07)	--
Trimester 1	2,483,224	290,768 (12)	58,967	7,246 (12)	1.03 (1.00, 1.07)	--	46,764	5,718 (12)	1.04 (1.00, 1.07)	--
Trimester 2	2,483,156	290,748 (12)	55,448	6,834 (12)	1.03 (1.00, 1.07)	--	50,351	6,150 (12)	1.04 (1.00, 1.07)	--
Trimester 3	2,475,357	290,367 (12)	64,045	7,858 (12)	1.03 (1.00, 1.07)	--	40,776	5,030 (12)	1.04 (1.01, 1.08)	--

Note: aOR, adjusted odds ratio; CI, confidence interval; BOE, barrel of oil equivalents of oil and gas; GT, greater than; EM, effect modification.

[a]Logistic regression models adjusted for inactive well count; child's sex, birth month and birth year; maternal education, age, race/ethnicity, Kotelchuck prenatal care index, parity; air basin, NO_2 concentration, and ICE for income.

Supplemental Table 1.7. Adjusted mean difference of term birth weight (grams) associated with oil and gas production volume by urban/rural status. Data for Figure 1.4D. Effect modification p-values were derived from two-sample z-tests using strata-specific estimates and variances; significance at α=0.05.

Prod volume categories	No BOE (ref)	1-100 BOE/day			GT 100 BOE/day		
	n	n	aDiff (95% CI)	EM p-value	n	aDiff (95% CI)	EM p-value
Rural[a]							
Entire pregnancy	297,643	8,339	3 (-11, 18)	0.62	1,590	-36 (-54, -17)	0.001
Trimester 1	297,772	8,205	4 (-10, 18)	0.47	1,595	-39 (-59, -19)	0.0003
Trimester 2	297,825	7,676	3 (-12, 18)	0.71	2,071	-27 (-45, -8)	0.01
Trimester 3	298,014	8,215	4 (-11, 20)	0.41	1,343	-30 (-48, -12)	0.001
Urban[a]							
Entire pregnancy	2,311,722	55,565	-5 (-10, 1)	--	43,770	1 (-5, 8)	--
Trimester 1	2,312,489	54,879	-5 (-11, 1)	--	43,689	2 (-4, 9)	--
Trimester 2	2,312,428	51,580	-5 (-11, 1)	--	47,049	2 (-4, 8)	--
Trimester 3	2,312,972	59,609	-6 (-12, 0)	--	38,476	5 (-2, 12)	--

Note: aDiff, adjusted mean difference (grams); CI, confidence interval; BOE, barrel of oil equivalents of oil and gas; GT, greater than;
EM, effect modification.
[a]Linear regression models adjusted for inactive well count; child's gestational age, sex, birth month and birth year; maternal education,
age, race/ethnicity, Kotelchuck prenatal care index, parity; air basin, NO_2 concentration, and ICE for income.

Supplemental Table 1.8. Rural adjusted odds ratios and mean difference (grams) for adverse birth outcomes associated with exposure to oil and gas production during the entire pregnancy by maternal race/ethnicity. Effect modification p-values were derived from two-sample z-tests using strata-specific estimates and variances; significance at α=0.05. Non-Hispanic Whites were used as the reference in z-tests. EM p-values were not reported for categories with no observations.

Prod volume categories	No BOE (ref)		1-100 BOE/day				GT 100 BOE/day			
	n	Cases (%)	*n*	Cases (%)	EE (95% CI)	EM p-value	n	Cases (%)	EE (95% CI)	EM p-value
Low birth weight[a]										
Asian/Pacific Islander	28,530	1,689 (6)	403	17 (4)	1.07 (0.59, 1.95)	1.00	326	23 (7)	2.04 (1.21, 3.43)	0.29
Black	6,313	494 (8)	81	4 (5)	0.65 (0.22, 1.96)	1.00	13	1 (8)	1.40 (0.47, 4.16)	1.00
Hispanic	164,739	7,735 (5)	5,828	283 (5)	1.16 (0.99, 1.37)	1.00	655	36 (6)	1.31 (1.02, 1.69)	1.00
Other	8,262	407 (5)	119	7 (6)	0.96 (0.47, 1.96)	1.00	27	0 (0)	--	--
White (ref)	110,644	4,126 (4)	2,526	89 (4)	1.04 (0.79, 1.37)	--	668	34 (5)	1.37 (0.89, 2.12)	--
Preterm birth[a]										
Asian/Pacific Islander	28,530	1,831 (6)	403	23 (6)	1.01 (0.56, 1.81)	1.00	326	17 (5)	1.06 (0.74, 1.52)	1.00
Black	6,313	590 (9)	81	2 (2)	--	0.56	13	0 (0)	--	--
Hispanic	164,739	11,882 (7)	5,828	458 (8)	1.14 (0.97, 1.33)	0.54	655	42 (6)	0.95 (0.69, 1.31)	1.00
Other	8,262	587 (7)	119	7 (6)	0.73 (0.21, 1.73)	1.00	27	1 (4)	0.53 (0.09, 3.16)	1.00
White (ref)	110,644	5,955 (5)	2,526	128 (5)	0.89 (0.70, 1.14)	--	668	39 (6)	0.94 (0.68, 1.31)	--
Small for gestational age[a]										
Asian/Pacific Islander	28,530	4,458 (16)	403	71 (18)	1.30 (0.90, 1.87)	1.00	326	51 (16)	1.09 (0.84, 1.42)	0.51
Black	6,313	978 (15)	81	12 (15)	0.81 (0.42, 1.57)	1.00	13	1 (8)	0.38 (0.12, 1.24)	1.00
Hispanic	164,739	17,307 (11)	5,828	646 (11)	1.08 (0.97, 1.22)	1.00	655	79 (12)	1.21 (0.91, 1.61)	1.00
Other	8,262	899 (11)	119	14 (12)	0.96 (0.50, 1.83)	1.00	27	6 (22)	2.03 (0.86, 4.58)	1.00
White (ref)	110,644	9,392 (9)	2,526	223 (9)	1.03 (0.84, 1.27)	--	668	74 (11)	1.24 (1.00, 1.56)	--
Term birth weight[b]										
Asian/Pacific Islander	26,699	--	380	--	-13 (-67, 42)	1.00	309	--	-23 (-62, 15)	1.00
Black	5,723	--	79	--	6 (-136, 148)	1.00	13	--	-115 (-257, 27)	1.00
Hispanic	152,857	--	5,370	--	-6 (-22, 10)	0.11	613	--	-23 (-53, 7)	1.00
Other	7,675	--	112	--	82 (-28, 192)	1.00	26	--	-147 (-320, 26)	1.00
White (ref)	104,689	--	2,398	--	23 (-4, 50)	--	629	--	-32 (-65, 0)	--

Note: EE, effect estimate; CI, confidence interval; BOE, barrel of oil equivalents of oil and gas; GT, greater than; EM, effect modification.

[a]Logistic regression models (odds ratio) adjusted for inactive well count; child's sex, birth month and birth year; maternal education, age, Kotelchuck prenatal care index, parity; air basin, NO_2 concentration, and ICE for income.

[b]Linear regression model (mean difference, grams) also adjusted for gestational age.

Supplemental Table 1.9. Urban adjusted odds ratios and mean difference (grams) for adverse birth outcomes associated with exposure to oil and gas production during the entire pregnancy by maternal race/ethnicity.

Prod volume categories	No BOE (ref)		1-100 BOE/day			GT 100 BOE/day		
	n	Cases (%)	*n*	Cases (%)	EE (95% CI)	n	Cases (%)	EE (95% CI)
Low birth weight[a]								
Asian/Pacific Islander	305,228	17,621 (6)	7,253	389 (5)	1.06 (0.94, 1.19)	6,182	376 (6)	1.10 (0.98, 1.23)
Black	136,701	13,034 (10)	3,655	370 (10)	1.13 (1.00, 1.29)	4,299	391 (9)	1.04 (0.93, 1.17)
Hispanic	1,417,139	71,208 (5)	34,986	1,791 (5)	1.01 (0.95, 1.07)	22,459	1,172 (5)	0.98 (0.92, 1.05)
Other	54,716	3,119 (6)	1,093	65 (6)	1.20 (0.87, 1.65)	1,209	70 (6)	1.18 (0.87, 1.61)
White	568,629	22,551 (4)	12,698	546 (4)	1.06 (0.96, 1.18)	12,708	452 (4)	0.87 (0.78, 0.98)
Preterm birth[a]								
Asian/Pacific Islander	305,228	18,550 (6)	7,253	388 (5)	1.05 (0.92, 1.20)	6,182	354 (6)	1.00 (0.87, 1.15)
Black	136,701	13,956 (10)	3,655	408 (11)	1.11 (0.97, 1.28)	4,299	398 (9)	0.91 (0.80, 1.04)
Hispanic	1,417,139	103,664 (7)	34,986	2,540 (7)	0.99 (0.94, 1.05)	22,459	1,630 (7)	0.93 (0.88, 0.99)
Other	54,716	3,854 (7)	1,093	88 (8)	1.15 (0.88, 1.50)	1,209	80 (7)	0.94 (0.70, 1.25)
White	568,629	30,667 (5)	12,698	696 (5)	0.99 (0.90, 1.10)	12,708	625 (5)	0.93 (0.84, 1.03)
Small for gestational age[a]								
Asian/Pacific Islander	305,228	48,463 (16)	7,253	1,185 (16)	1.04 (0.95, 1.13)	6,182	1,045 (17)	1.11 (1.03, 1.21)
Black	136,701	24,563 (18)	3,655	702 (19)	1.14 (1.05, 1.25)	4,299	770 (18)	1.11 (1.00, 1.22)
Hispanic	1,417,139	160,491 (11)	34,986	4,149 (12)	1.01 (0.97, 1.06)	22,459	2,558 (11)	0.99 (0.94, 1.04)
Other	54,716	6,581 (12)	1,093	148 (14)	1.31 (1.07, 1.59)	1,209	147 (12)	1.17 (0.95, 1.15)
White	568,629	50,556 (9)	12,698	1,155 (9)	0.98 (0.91, 1.05)	12,708	1,219 (10)	1.01 (0.94, 1.09)
Term birth weight[b]								
Asian/Pacific Islander	286,678	--	6,865	--	-3 (-20, 14)	5,828	--	-16 (-30, -1)
Black	122,745	--	3,247	--	-35 (-53, -18)	3,901	--	-4 (-22, 14)
Hispanic	1,313,475	--	32,446	--	0 (-8, 7)	20,829	--	10 (1, 18)
Other	50,862	--	1,005	--	-23 (-55, 9)	1,129	--	-5 (-39, 29)
White	537,962	--	12,002	--	-2 (-12, 9)	12,083	--	3 (-8, 15)

Note: EE, effect estimate; CI, confidence interval; BOE, barrel of oil equivalents of oil and gas; GT, greater than.

[a]Logistic regression models (odds ratio) adjusted for inactive well count; child's sex, birth month and birth year; maternal education, age, Kotelchuck prenatal care index, parity; air basin, NO_2 concentration, and ICE for income.

[b]Linear regression model (mean difference) also adjusted for gestational age.

Supplemental Table 1.10. Rural adjusted odds ratios and mean difference (grams) for adverse birth outcomes associated with exposure to oil and gas production during the entire pregnancy by air basin. Effect modification p-values were derived from two-sample z-tests using strata-specific estimates and variances; significance at α=0.05. Sacramento Valley was used the reference for all z-tests.

Prod volume categories	No BOE (ref)		1-100 BOE/day				GT 100 BOE/day			
	n	Cases (%)	*n*	Cases (%)	EE (95% CI)	EM p-value	n	Cases (%)	EE (95% CI)	EM p-value
Low birth weight[a]										
Sacramento Valley (ref)	51,150	2,132 (4)	1,036	38 (4)	0.84 (0.52, 1.35)	--	237	14 (6)	1.37 (0.82, 2.26)	--
San Joaquin Valley	157,748	7,509 (5)	4,501	214 (5)	1.14 (0.95, 1.37)	1.00	571	37 (6)	1.48 (1.05, 2.07)	1.00
South Central Coast	36,860	1,552 (4)	2,820	120 (4)	1.05 (0.84, 1.30)	1.00	300	14 (5)	1.32 (0.90, 1.93)	1.00
South Coast	72,730	3,258 (4)	600	28 (5)	1.38 (0.86, 2.23)	0.97	581	29 (5)	1.41 (1.14, 1.74)	1.00
Preterm birth[a]										
Sacramento Valley (ref)	51,150	2,999 (6)	1,036	56 (5)	0.68 (0.47, 0.99)	--	237	16 (7)	0.86 (0.56, 1.32)	--
San Joaquin Valley	157,748	11,575 (7)	4,501	344 (8)	1.02 (0.86, 1.21)	0.45	571	48 (8)	1.05 (0.82, 1.33)	1.00
South Central Coast	36,860	2,052 (6)	2,820	183 (7)	1.09 (0.88, 1.35)	0.17	300	8 (3)	0.54 (0.33, 0.87)	0.67
South Coast	72,730	4,219 (6)	600	35 (6)	1.21 (0.91, 1.61)	0.14	581	27 (5)	0.98 (0.75, 1.30)	1.00
Small for gestational age[a]										
Sacramento Valley (ref)	51,150	4,769 (9)	1,036	89 (9)	1.03 (0.76, 1.40)	--	237	16 (7)	0.80 (0.54, 1.18)	--
San Joaquin Valley	157,748	16,910 (11)	4,501	498 (11)	1.09 (0.94, 1.27)	1.00	571	87 (15)	1.53 (1.19, 1.97)	0.05
South Central Coast	36,860	3,539 (10)	2,820	296 (11)	1.03 (0.90, 1.20)	1.00	300	38 (13)	1.38 (1.22, 1.57)	0.11
South Coast	72,730	7,816 (11)	600	83 (14)	1.26 (1.03, 1.53)	1.00	581	70 (12)	1.00 (0.81, 1.25)	1.00
Term birth weight[b]										
Sacramento Valley (ref)	48,151	--	980	--	18 (-27, 64)	--	221	--	-48 (-100, 4)	--
San Joaquin Valley	146,173	--	4,157	--	-1 (-23, 21)	1.00	523	--	-38 (-77, 1)	1.00
South Central Coast	34,808	--	2,637	--	2 (-22, 25)	1.00	292	--	-42 (-74, -10)	1.00
South Coast	68,511	--	565	--	-7 (-30, 17)	1.00	554	--	-31 (-48, -14)	1.00

Note: EE, effect estimate; CI, confidence interval; BOE, barrel of oil equivalents of oil and gas; GT, greater than; EM, effect modification.

[a]Logistic regression models (odds ratio) adjusted for inactive well count; child's sex, birth month and birth year; maternal education, age, race/ethnicity, Kotelchuck prenatal care index, parity; NO_2 concentration and ICE for income.

[b]Linear regression model (mean difference, grams) also adjusted for gestational age.

Supplemental Table 1.11. Urban adjusted odds ratios and mean difference (grams) for adverse birth outcomes associated with exposure to oil and gas production during the entire pregnancy by air basin.

Prod volume categories	No BOE (ref)		1-100 BOE/day			GT 100 BOE/day		
	n	Cases (%)	*n*	Cases (%)	EE (95% CI)	n	Cases (%)	EE (95% CI)
Low birth weight[a]								
Sacramento Valley	238,174	11,524 (5)	0	0 (0)	--	0	0 (0)	--
San Joaquin Valley	377,626	20,833 (6)	9,717	511 (5)	1.02 (0.91, 1.13)	1,517	76 (5)	0.95 (0.78, 1.16)
South Central Coast	131,912	5,886 (4)	1,076	37 (3)	0.70 (0.54, 0.93)	288	15 (5)	1.26 (0.96, 1.65)
South Coast	1,734,701	89,290 (5)	48,892	2,613 (5)	1.05 (1.00, 1.11)	45,052	2,370 (5)	1.00 (0.95, 1.05)
Preterm birth[a]								
Sacramento Valley	238,174	14,318 (6)	0	0 (0)	--	0	0 (0)	--
San Joaquin Valley	377,626	30,907 (8)	9,717	803 (8)	0.99 (0.90, 1.10)	1,517	129 (9)	0.98 (0.84, 1.14)
South Central Coast	131,912	7,730 (6)	1,076	64 (6)	0.86 (0.64, 1.17)	288	23 (8)	1.46 (1.17, 1.82)
South Coast	1,734,701	117,736 (7)	48,892	3,253 (7)	1.01 (0.96, 1.06)	45,052	2,935 (7)	0.97 (0.92, 1.02)
Small for gestational age[a]								
Sacramento Valley	238,174	24,960 (10)	0	0 (0)	--	0	0 (0)	--
San Joaquin Valley	377,626	44,604 (12)	9,717	1,131 (12)	1.00 (0.90, 1.10)	1,517	158 (10)	0.92 (0.80, 1.05)
South Central Coast	131,912	13,798 (10)	1,076	120 (11)	1.06 (0.90, 1.25)	288	34 (12)	1.20 (0.85, 1.71)
South Coast	1,734,701	207,292 (12)	48,892	6,088 (12)	1.04 (1.00, 1.08)	45,052	5,547 (12)	1.04 (1.01, 1.080
Term birth weight[b]								
Sacramento Valley	223,856	--	0	--	--	0	--	--
San Joaquin Valley	346,719	--	8,914	--	2 (-10, 14)	1,388	--	2 (-16, 20)
South Central Coast	124,182	--	1,012	--	4 (-30, 37)	265	--	-52 (-95, -8)
South Coast	1,616,965	--	45,639	--	-6 (-13, 1)	42,117	--	1 (-5, 8)

Note: EE, effect estimate; CI, confidence interval; BOE, barrel of oil equivalents of oil and gas; GT, greater than.

[a]Logistic regression models (odds ratio) adjusted for inactive well count; child's sex, birth month and birth year; maternal education, age, race/ethnicity, Kotelchuck prenatal care index, parity; NO_2 concentration and ICE for income.

[b]Linear regression model (mean difference, grams) also adjusted for gestational age.

Supplemental Table 1.12. Adjusted odds ratios and mean difference (grams) for adverse birth outcomes associated with exposure to oil and gas production during the entire pregnancy by urban/rural status for sensitivity analysis models including maternal pre-pregnancy BMI and smoking during pregnancy (2007-2015). BMI and smoking were not available for 2006 births. Effect estimates did not change by >10% compared to main models. Main models excluded these two covariates in order to maximize sample size.

Prod volume categories	No BOE (ref)		1-100 BOE/day			GT 100 BOE/day		
	n	Cases (%)	*n*	Cases (%)	EE (95% CI)	n	Cases (%)	EE (95% CI)
Low birth weight[a]								
Rural	283,881	12,911 (5)	7,879	357 (5)	1.16 (1.00, 1.36)	1,477	81 (5)	1.32 (1.10, 1.59)
Urban	2,179,247	111,604 (5)	51,676	2,747 (5)	1.04 (0.99, 1.10)	40,147	2,062 (5)	0.96 (0.91, 1.01)
Preterm birth[a]								
Rural	283,881	18,236 (6)	7,879	521 (7)	1.05 (0.91, 1.20)	1,477	85 (6)	0.96 (0.77, 1.21)
Urban	2,179,247	146,242 (7)	51,676	3,439 (7)	1.01 (0.97, 1.06)	40,147	2,535 (6)	0.95 (0.90, 1.00)
Small for gestational age[a]								
Rural	283,881	29,451 (10)	7,879	862 (11)	1.11 (0.99, 1.24)	1,477	190 (13)	1.17 (0.97, 1.42)
Urban	2,179,247	254,729 (12)	51,676	6,361 (12)	1.05 (1.01, 1.09)	40,147	4,828 (12)	1.03 (0.99, 1.06)
Term birth weight[b]								
Rural	265,645	--	7,358	--	-2 (-17, 12)	1,392	--	-31 (-51, -11)
Urban	2,033,005	--	48,237	--	-5 (-11, 2)	37,612	--	3 (-3, 9)

Note: EE, effect estimate; CI, confidence interval; BOE, barrel of oil equivalents of oil and gas; GT, greater than.

[a]Logistic regression models (odds ratio) adjusted for inactive well count; child's sex, birth month and birth year; maternal BMI, smoking, education, age, race/ethnicity, Kotelchuck prenatal care index, parity; air basin, NO_2 concentration, and ICE for income.

[b]Linear regression model (mean difference) also adjusted for gestational age.

Supplemental Table 1.13. Adjusted odds ratios and mean difference (grams) for adverse birth outcomes associated with exposure to oil and gas production during the entire pregnancy by urban/rural status for sensitivity analysis models including an indicator for exposure to TRI facilities within 1 km. The variable was missing for 79,371 observations (3%). Effect estimates did not change by >10%, compared to main models.

Prod volume categories	No BOE (ref)		1-100 BOE/day			GT 100 BOE/day		
	n	Cases (%)	*n*	Cases (%)	EE (95% CI)	n	Cases (%)	EE (95% CI)
Low birth weight[a]								
Rural	318,488	14,451 (5)	8,957	400 (4)	1.11 (0.97, 1.28)	1,689	94 (6)	1.40 (1.14, 1.71)
Urban	2,482,413	127,533 (5)	59,685	3,161 (5)	1.04 (1.00, 1.09)	46,857	2,461 (5)	0.99 (0.95, 1.04)
Preterm birth[a]								
Rural	318,488	20,845 (7)	8,957	618 (7)	1.03 (0.91, 1.17)	1,689	99 (6)	0.97 (0.78, 1.21)
Urban	2,482,413	170,691 (7)	59,685	4,120 (7)	1.01 (0.97, 1.06)	46,857	3,087 (7)	0.95 (0.90, 1.00)
Small for gestational age[a]								
Rural	318,488	33,034 (10)	8,957	966 (11)	1.08 (0.97, 1.19)	1,689	211 (12)	1.22 (1.02, 1.45)
Urban	2,482,413	290,654 (12)	59,685	7,339 (12)	1.03 (1.00, 1.07)	46,857	5,739 (12)	1.04 (1.01, 1.07)
Term birth weight[b]								
Rural	297,643	--	8,339	--	3 (-11, 18)	1,590	--	-36 (-54, -17)
Urban	2,311,722	--	55,565	--	-5 (-10, 1)	43,770	--	1 (-5, 8)

Note: EE, effect estimate; CI, confidence interval; BOE, barrel of oil equivalents of oil and gas; GT, greater than.

[a]Logistic regression models (odds ratio) adjusted for inactive well count; child's sex, birth month and birth year; maternal education, age, race/ethnicity, Kotelchuck prenatal care index, parity; air basin, NO_2 concentration, ICE for income and TRI facilities indicator.

[b]Linear regression model (mean difference) also adjusted for gestational age.

Chapter 2: Residential proximity to hydraulically fractured oil and gas wells and adverse birth outcomes in urban and rural communities in California (2006-2015)

Publication: Tran K, Casey JA, Cushing LJ, Morello-Frosch R (2021) Residential proximity to hydraulically fractured oil and gas wells and adverse birth outcomes in urban and rural communities in California (2006-2015). Environmental Epidemiology, Forthcoming.

2.1 Abstract

Background: Prenatal exposure to hydraulic fracturing (HF), a chemically intensive oil and gas extraction method, may be associated with adverse birth outcomes, but no health studies have been conducted in California.
Objective: To assess the relationship between prenatal exposure to HF wells and perinatal outcomes in Kern and Los Angeles (LA) counties, where HF predominantly occurs.
Methods: We conducted a retrospective cohort study of 979,961 births to mothers in eight California counties with HF between 2006 and 2015. Exposed individuals had at least 1 well hydraulically fractured within 1 km of their residence during pregnancy; the reference population had no wells within 1 km, but at least one oil/gas well within 10 km. We examined associations between HF and low birth weight (LBW), preterm birth (PTB), small for gestational age birth (SGA), and term birth weight (tBW) using generalized estimating equations and assessing urban-rural effect modification in stratified models.
Results: Fewer than 1% of mothers (N=1,192) were exposed to HF during pregnancy. Among rural mothers, HF exposure was associated with increased odds of LBW (odds ratio [OR] = 1.74 and 95% confidence interval [CI]: 1.10, 2.75), SGA (OR = 1.68, 95% CI: 1.42, 2.27) and PTB (OR = 1.17, 95% CI: 0.64, 2.12), and lower tBW (mean difference: -73 g, 95% CI: -131, -15). Among urban mothers, HF exposure was positively associated with SGA (OR = 1.23, 95% CI: 0.98, 1.55), inversely associated with LBW (OR: 0.83, 95% CI: 0.63, 1.07) and PTB (OR: 0.65, 95% CI: 0.48, 0.87), and not associated with tBW (mean difference: -2 g, 95% CI: -35, 31).
Conclusions: HF proximity was associated with adverse birth outcomes, particularly among rural Californians.

2.2 Background

California is among the top-10 oil and top-15 natural gas producing U.S. states (US EIA 2020). Hydraulic fracturing (HF) is a common well stimulation technique for enhanced oil and gas recovery (Long et al. 2015c) and accounts for about 20% of California's oil and gas production (Long et al. 2015a). Uniquely, HF involves injecting water, proppants and chemicals into wells at high pressure to create cracks in rock formations, which maximizes extraction flow (Long et al. 2015c, 2015a). HF primarily occurs in California's Central Valley region, and compared to other states, most HF wells are shallower, more vertical, require less water per well, and use more concentrated chemical mixtures to recover primarily oil (Jackson et al. 2015; Long et al. 2015c, 2015a; US EPA 2015a). Although chemicals make up about 1% or less of the mixture, and HF usually takes less than a day, these chemicals may pose potential health hazards (Long et al. 2015a).

One exposure pathway is via contamination of surface or groundwater with wastewater associated with HF (flowback) and oil and gas production (produced water) (Long et al. 2015a). Compared to conventional non-HF extraction, HF produces greater volumes of wastewater, which can include fugitive oil and gas, salts, organic and inorganic chemicals, radioactive material, and additives that can react with one another to generate byproducts–via flowback (Long et al. 2015a). In California, between January 2011 and June 2014, nearly 60% (or 720,000 m3) of wastewater generated from stimulated wells was disposed in unlined pits for evaporation and percolation while about a quarter of the wastewater was injected (Long et al. 2015a). In the process of injecting wastewater into wells, accidental spills during transfer and transport, and leaks in storage wells can release contaminants into the environment. The highest number of wastewater-related spills across California were recorded in Kern County (Central Valley) between 2009 and 2014 (Long et al. 2015a). In Pennsylvania, trace metals related to HF (e.g., barium, strontium) have been found in private well-water (Caron-Beaudoin et al. 2021). In California, concerns about health and environmental impacts of water contamination associated with HF resulted in passage of Senate Bill 4 (SB4) in 2014 (Pavley 2013) requiring oil and gas companies to expand monitoring and disclose chemicals used during fracking.

HF chemicals could also affect public health via air pollution emitted during well drilling, handling and mixing of chemicals for injection, hydraulic fracturing, and management of recovered fluids and waste products (McKenzie et al. 2012b; Shonkoff et al. 2019). Volatile organic compounds (VOCs), such as benzene, toluene, ethylbenzene, and xylene (BTEX) and formaldehyde, have been the most commonly measured pollutants in and near HF wells and may be associated with adverse birth outcomes (Bolden et al. 2015; Caron-Beaudoin et al. 2018; Chang et al. 2017; Maroziene and Grazuleviciene 2002). Measured emissions during drilling, hydraulic fracturing, flowback and production at 5-10 well pads showed that emission rates of benzene and most VOCs were highest during flowback (Collett and Colorado State University 2016; Hecobian et al. 2019). In several regions with intense HF activity, higher concentrations of VOCs have also been measured in ambient air compared to regions without HF (Caron-Beaudoin et al. 2021).

Pregnancy is a vulnerable period of human development, and adverse birth outcomes are primary predictors of infant mortality and morbidity (Bhutta et al. 2002; Hack et al. 1995; Moster et al.

2008; Saigal and Doyle 2008). Studies indicate associations between prenatal exposure to oil and gas development (OGD) activities (HF [most studies] and conventional extraction methods) and reductions in birth weight (tBW) (Caron-Beaudoin et al. 2021; Hill 2018; Stacy et al. 2015), increased odds or incidence of low birth weight (LBW) (Currie et al. 2017; Hill 2018), preterm birth (PTB) (Caron-Beaudoin et al. 2021; Casey et al. 2015c; Cushing et al. 2020; Gonzalez et al. 2020a; Walker Whitworth et al. 2018; Whitworth et al. 2017), and small for gestational age birth (SGA) (Hill 2018; Stacy et al. 2015). Statistically insignificant (Caron-Beaudoin et al. 2021; Casey et al. 2015c; Whitworth et al. 2017) or inverse associations (McKenzie et al. 2014; Stacy et al. 2015) for some birth outcomes have also been observed. Our previous California study found exposure to all OGD (mostly not involving HF) was associated with decreased tBW and increased odds of LBW and SGA in rural areas, and increased odds of SGA in urban areas (Tran et al. 2020a). Because unique elements of HF, including the use of additional chemicals and large volumes of wastewater generated, may pose additional health risks beyond risks from conventional extraction, which we previously analyzed (Tran et al. 2020a), we extend our work to examine associations between prenatal exposure to HF and four birth outcomes (tBW, LBW, PTB, and SGA) by focusing on those California regions where HF is prevalent.

2.3 Methods

Study population

The study population, previously described (Tran et al. 2020a), consisted of births between January 1, 2006 and December 31, 2015 derived from the California Department of Public Health (CDPH) birth records. The dataset included maternal and infant characteristics such as self-reported race/ethnicity and infant sex, and maternal residential addresses were geocoded with ArcGIS 10.6 (Esri, Redlands, CA). From all 2006-2015 births (5.2 million), we limited our analysis to births in four air basins (Sacramento Valley, San Joaquin Valley, South Central Coast and South Coast with 26 counties) where most of California's oil and gas extraction activities occur after excluding births with missing data and birth defects (**Supplemental Figure 2.1**). Mothers also had to reside within 10 km of at least one well, a criterion applied to limit unmeasured confounding and enhance comparability of the exposed and unexposed populations (Tran et al. 2020a). For this analysis, we further limited our cohort to births from 8 counties (Colusa, Fresno, Glenn, Kern, Los Angeles, Orange, Santa Barbara and Ventura) with at least one maternal residence within 1 km of at least one HF well during the study period. After removing births with missing data, the study population consisted of 979,961 live births. Ninety percent of HF wells were in Kern County (**Figure 2.1**) and 4% in Los Angeles County. Study protocols were approved by the Institutional Review Boards of the CDPH (#13-05-1231) and the University of California, Berkeley (# 2013-10-5693).

Birth outcomes

We assessed the relationship between HF and four birth outcomes: 1) continuous tBW (grams (g), among births at ≥37 completed weeks), 2) LBW (<2500 grams), 3) PTB (<37 completed weeks), and SGA birth (birth weight less than the US sex-specific 10th percentile of weight for each week of gestation) (Talge et al. 2014b). Gestational age was estimated by subtracting the last menstrual period (LMP) date from the date of birth.

Exposure assessment
We derived data on confirmed HF wells from two sources: 1) the California Council on Science and Technology's (CCST) well stimulation report (Vol 1, Appendix M, hereafter CCST report) (CCST 2015) and 2) California Division of Oil, Gas and Geothermal Resources' (DOGGR, now CalGEM) well stimulation treatment (i.e., HF, acid fracturing and matrix acidization) disclosure database (CA DOGGR 2019). Both datasets contain unique American Petroleum Institute (API) numbers for each well, latitude, longitude, and approximate HF dates. We compiled HF records from the CCST report for January 2005 to December 2013 and the remaining HF well records for January 2014 to December 2015 from DOGGR.

CCST's methodology for compiling Appendix M is described in detail elsewhere (Volume 1, Appendix I (Long et al. 2015c)). Briefly, HF wells were identified by reviewing OGD permit records and scanning for "frac " for evidence of HF as there was no systematic reporting requirement prior to SB4 in 2014. Due to the large number of well records in Kern County, CCST randomly sampled and reviewed 20% of records while for Los Angeles County, they reviewed 80% of records that were made available by county officials; 100% of records were reviewed in all other counties. CCST extracted approximate HF dates from permits or other sources such as regional air or water districts (CCST 2015). The DOGGR stimulation disclosure database was initiated in January 2014 with the adoption of California SB4. We filtered on HF, the bulk of the permit records, among the three types of stimulation techniques. There was a sharp decrease in the number of HF wells in 2014 (reason unknown) while operators adjusted to SB4 implementation (Long et al. 2015b).

After compiling confirmed HF wells from the two data sources, we used the stimulation date to identify whether HF occurred during each pregnancy and the well location to identify proximity to residences. We then summed the number of HF events within 1 km of each mother's residence for each month of pregnancy using R version 3.3.1 (R Development Core Team, Auckland, New Zealand). HF wells that were not stimulated during a woman's pregnancy period did not contribute to exposure. We classified women who had at least one well stimulated within 1 km of their residential address at any point during pregnancy as exposed; prior literature found strongest associations with health indicators and exposure to OGD within this radius (Boyle et al. 2017; McKenzie et al. 2012a; Meng 2015a; Walker Whitworth et al. 2018; Whitworth et al. 2017). Women without any oil or gas wells within 1 km, but at least one well (whether HF or not during pregnancy) within 10 km, were classified as unexposed (**Figure 2.2**).

Covariates
To address potential confounding, our models controlled for several individual-level maternal characteristics and area-level variables. Individual-level covariates from birth records were identified a priori as potential confounders based on prior studies. Infant covariates included sex (male/female), month and year of conception based on the date of LMP (both categorical) to control for seasonal and secular trends. Maternal covariates included age (< 20, 20-24, 25-29, 30-34, 35+), self-reported race and ethnicity (non-Hispanic White, Black, Asian-Pacific Islander (API), Other and Hispanic), educational attainment (<high school, high school graduate/GED, some college, college+), Kotelchuck index of prenatal care (Alexander and Kotelchuck 1996; Kotelchuck 1994), and parity (nulliparous vs. multiparous). We aggregated Asian subgroups into the API category, and other racial/ethnic groups with small sample sizes into the Other category

to ensure adequate subgroup sample size. In the tBW model, we also added mean-centered and mean-centered squared gestational age (continuous) to allow for nonlinearity. Though mothers' smoking status during pregnancy and pre-pregnancy body mass index (BMI) are known predictors for adverse birth outcomes, they were not included because these variables were not available for 2006 births and our previous sensitivity analyses (Tran et al. 2020a) indicated that including them when available did not substantially change effect estimates.

Area-level variables consisted of California Air Resources Board designated air basins, census-tract based urban-rural classification (urban tract if at least 60% of its area overlapped with an urbanized or urban area as defined by the US Census Bureau (US Census Bureau), rural otherwise), modeled annual average nitrogen dioxide (NO_2) concentrations (Kim et al. 2020a) as a proxy for traffic-related air pollution (Brook et al. 2007; Kendrick et al. 2015), and Index of Concentration at the Extremes (ICE) (quartiles), a measure of neighborhood-level relative deprivation or affluence based on household income by census tract (Massey 1996). ICE for income was categorized into quartiles and ranged between 1 (concentration of affluence) and -1 (concentration of deprivation). The variable reflects the difference between the number of people with median household income in the top 80^{th} percentile and the number of people with median household income in the lower 20th percentile within census tracts (urban tracts) or county (rural tracts), adjusted by the total tract/county population. These covariates were included to account for neighborhood and regional differences in air quality, economic activity, and sources of emissions (Arruti et al. 2011; Finkelstein et al. 2003; O'Neill et al. 2003; Wunderli and Gehrig 1990; Zhao et al. 2009).

Statistical analyses

Statistical analyses were conducted in SAS 9.4 (SAS Institute, Cary, NC). We constructed separate models for each of our four birth outcomes to assess the association between prenatal exposure to HF and odds of PTB, LBW or SGA, or mean tBW. We used generalized estimating equations to account for clustering within census tracts. For the primary analysis, we compared births to mothers who were exposed to HF to those who were not exposed to any OGD during pregnancy within 1 km. As our previous study revealed significant effect modification (EM) by urbanicity (Tran et al. 2020a), we stratified models by urban and rural tracts (Model 1). We then tested for significant heterogeneity between strata-specific estimates by modeling urbanicity as an interaction term to derive p-values for two-sample z-tests using model-estimated beta coefficients and variances (Buckley et al. 2017; UCLA: Statistical Consulting Group). Due to the small exposed sample size, we evaluated model overadjustment by adjusting for one maternal covariate at a time and comparing the effect estimates between the fully adjusted models and single-covariate adjusted models.

Sensitivity analysis

We conducted a sensitivity analysis including broader exposure reference groups: mothers with no wells of any type and mothers with active or inactive wells that were not identified as HF wells within 1 km (**Figure 2.2**). Because HF and conventional wells are often clustered, which may confound associations between HF exposure and adverse birth outcomes, the sensitivity analysis adjusted for exposure to non-HF active and inactive wells (Model 2). The number of inactive wells was categorized as 0, 1, 2-5, ≥ 6 following Tran et al. 2020. Production volume was calculated as the sum of total monthly barrels of oil equivalent (BOE) from oil and gas wells during pregnancy (normalized by length of pregnancy and categorized as 0, 1-100 or >100 BOE/day) (Tran et al. 2020a).

2.4 Results

The study population consisted of 979,961 births to mothers residing within 10 km of an oil or gas well between January 2006 and December 2015 in the eight California counties (**Supplemental Figure 2.1**). Of these, 0.1% (n = 1,162) were exposed to HF in utero (**Figure 2.2**). Mean birth weight was 3,310 grams (standard deviation = 523) (**Table 2.1**). Five percent (n = 52,378) of all births were LBW, 7% (n = 70,772) preterm, and 12% (n = 120,590) SGA. PTB was 2% higher among the reference group compared to the HF exposed while SGA was 4% higher among the exposed group. HF exposed mothers, on average, were exposed to 2 HF wells within 1 km and a maximum of 20 HF wells (**Table 2.1**). Exposed mothers were also more educated [31% vs. 24% college or more educated], older [29% vs. 26% ages 30-34], more often non-Hispanic Black [16% vs. 4%], more likely to have inadequate prenatal care [13% vs. 10%], and more likely to not have previously given birth [42% vs. 39% nulliparous]. Relative to unexposed mothers, exposed mothers were more likely to reside in Kern [17% vs. 3%], Los Angeles [65% vs. 57%] and Ventura counties [10% vs. 3%], rural areas [20% vs. 7%] and economically segregated areas [e.g., 35% vs. 25% in neighborhoods with concentrated poverty and 37% vs. 25% with concentrated affluence].

In overall unstratified models, effect estimates showed positive associations between prenatal exposure to HF wells and SGA and reduced tBW as well as inverse associations between exposure and LBW and PTB (**Supplemental Table 2.1**). **Table 2.2** shows our models stratified by urbanicity. When fully adjusted, the associations differed by urban and rural tracts (**Table 2.2**); EM p-values were 0.007, 0.09, 0.10 and 0.05 for LBW, PTB, SGA and tBW, respectively. Among rural mothers, exposure to HF wells was associated with increased odds for LBW (OR = 1.74, 95% CI: 1.10, 2.75), PTB (OR = 1.17, 95% CI: 0.64, 2.12) and SGA (OR = 1.68, 95% CI: 1.42, 2.27) and decreased tBW (mean difference = -73 grams, 95% CI: -131, -15) (Table 2). Among urban mothers, HF exposure was associated with increased odds of SGA (OR =1.23, 95% CI: 0.98, 1.55), but not with tBW (mean difference = -2, 95% CI: -35, 31), as well as reduced odds of PTB (OR = 0.65, 95% CI: 0.48, 0.87) and LBW (OR = 0.83, 95% CI: 0.63, 1.07). Compared to the single maternal covariate adjusted models (**Supplemental Table 2.2**), results were qualitatively similar, albeit attenuated.

In our sensitivity analysis with an expanded reference population (no wells of any type within 1 km as well as non-HF OGD wells within 1km), results were qualitatively similar to those from the primary analysis for all four birth outcomes (**Supplemental Tables 2.1-2.3**). However, evidence of urban-rural effect modification was weaker in the sensitivity analysis. Except for LBW and tBW among the rural population, most effect estimates did not change by >10%.

2.5 Discussion

To our knowledge, this study is the first to examine the association between prenatal exposures to HF and adverse birth outcomes in California. We found that prenatal exposure to HF was associated with all four adverse birth outcomes among rural residents, with the strongest associations observed for LBW, SGA and tBW. While the direction of the urban effect estimate was consistent with the rural communities for SGA, we observed inverse associations for PTB and LBW and no association with tBW. In our evaluation of overadjustment, the effect estimates remained stable.

Results remained consistent, with slightly weaker associations, in both rural and urban tracts in our sensitivity analysis including a larger reference population. With a broader definition for the unexposed group, there is a higher likelihood of exposure misclassification as 80% of Kern–where the majority of HF occurs–well records were not reviewed to confirm HF status; this may have led to the observed weaker associations. Nevertheless, the consistency of results between the primary and sensitivity analyses suggests that HF exposure may influence birth outcomes independent of the presence of conventional wells.

Similar to our previous analysis of exposure to all OGD (Tran et al. 2020a), we observed differences in effect estimates between rural and urban areas. The significant EM p-values for LBW and tBW suggest that urbanicity modifies the association between HF exposure and birth weight. This may occur because urban regions tend to have more diverse mobile and stationary sources of ambient air pollution, and OGD likely contributes relatively less to urban ambient air pollution, making detection of the unique effects from OGD, and HF in particular, more challenging. Rural residents are also more likely to rely on groundwater sources for their drinking water, which may more likely be untreated if contaminated by OGD-related chemicals (Balazs and Ray 2014). Most HF wells in Kern County are located in relatively shallow reservoirs, where groundwater protected for drinking water might be found within a few hundred feet (Long et al. 2015b).

Our findings were consistent with those of previous studies that examined exposure to HF in rural and urban Pennsylvania and urban Texas. Evidence of a relationship between HF and LBW has been sparse; two studies observed increased risk of LBW associated with HF exposure in Pennsylvania (Currie et al. 2017; Hill 2018). Evidence of associations between HF exposure and tBW has been mixed; among five studies, two found no relationship (Pennsylvania, Texas) (Casey et al. 2015c; Whitworth et al. 2017), and three found decreased tBW in Pennsylvania (Currie et al. 2017; Hill 2018; Stacy et al. 2015). Cohort studies in Pennsylvania and Texas suggested that prenatal exposure to HF significantly increased odds of PTB by 14% to 100% (Casey et al. 2015c; Walker Whitworth et al. 2018; Whitworth et al. 2017). We observed a PTB estimate similar in magnitude and direction to those findings among the rural population, while the association was inverse in urban areas. Among the three studies that evaluated SGA, two studies (Pennsylvania, Texas) found no association (Casey et al. 2015c; Whitworth et al. 2017) while the other Pennsylvania study observed a similar magnitude of increased odds of SGA as in our study (Stacy et al. 2015). The observed differences across studies may be partially explained by differences in exposure sources, setting, and OGD infrastructure. Ambient air pollution levels and pollution sources in rural California may be more similar to those of rural Pennsylvania than those observed in urban Californian communities. New well pad development, drilling of new wells and horizontal or directional drilling also occur less frequently in California compared to Pennsylvania and Texas where infrastructure is less mature and wells are deeper, meaning higher volumes of water are pumped into wells and collected as flowback (Long et al. 2015a).

Additionally, California primarily produces oil (Long et al. 2015c) while Pennsylvania mainly produces gas and Texas produces mainly gas in the northern region. The constituents of fracking fluid vary by region based on hydrocarbon properties (e.g., oil is more viscous than gas) and local geology (Long et al. 2015c), meaning the type and concentration of chemicals that may contaminate air and waterways likely also vary by region.

Associations between exposure to HF and SGA were stronger than those we previously observed in California for exposure to high production volume from mostly conventional wells in both rural [OR = 1.22 (95% CI: 1.02, 1.45)] and urban [OR = 1.04 (95% CI: 1.01, 1.07)] areas (Tran et al. 2020a), This suggests that HF treatment may present additional hazards or enhanced health risks compared to conventional OGD operations. However, because only a small proportion of births were exposed to HF (<0.01% of births to mothers residing within 10 km of any well in the 8 counties), the risk difference between the exposed and unexposed is smaller compared to that for exposure to all types of actively producing wells (which affected a larger population, 4% of California births to mothers residing within 10 km of any well in 23 counties). Within 1 km, HF wells likely contribute a sizeable proportion of OGD-related air pollution. Truck traffic required to transport materials and equipment to and from the well pad for HF (Long et al. 2015c) is likely a primary source. HF in California typically requires about 100-200 diesel truck trips per vertical well, and 200-400 trips per horizontal well (Long et al. 2015a). Ambient $PM_{2.5}$, a component of diesel particulates, has been associated with higher odds of SGA (Gray et al. 2014; Hyder et al. 2014; Zhu et al. 2015). Air samples collected in five states (Arizona, Ohio, Wyoming, Colorado, and Pennsylvania) near stimulated well sites and wastewater impoundments from distances as close as 27-320 meters of unconventional OGD sites revealed elevated levels of VOCs, including BTEX (Macey et al. 2014). Benzene from unconventional wells has been measured at elevated levels within 1 km from oil and gas fields in several states (Halliday et al. 2016; Macey et al. 2014; Maskrey et al. 2016; McKenzie et al. 2012b; Rich and Orimoloye 2016; Swarthout et al. 2013; Thompson et al. 2014). This indicates that OGD equipment and volatized chemicals from percolation pits can contribute to OGD emissions. VOCs and BTEX may be associated with decreased birth weight (Bolden et al. 2015; Chang et al. 2017) and substantial decreases in birth weight can result in SGA. BTEX is not only found in emissions but also in groundwater samples after spills at HF sites (Gross et al. 2013a). As water contamination risks are not well understood, current water treatment practices may not prevent exposure to HF-related chemicals.

Besides significant associations with SGA, exposure to HF was also unexpectedly inversely associated with PTB and LBW within urban areas. Among studies that evaluated birth outcomes and unconventional OGD, one revealed an inverse association with exposure to HF and PTB in Pennsylvania (Stacy et al. 2015). Decreased odds of PTB have also been observed with increasing levels of ambient air pollution (Jalaludin et al. 2007; Stieb et al. 2012). The inverse association between HF and PTB observed in our study may be due to residual confounding from area-level SES characteristics or environmental factors that we could not account for in our analyses. Additionally, live birth bias can result from the depletion of susceptibles, which may occur if exposed compared to unexposed mothers were more likely to experience fetal loss (Bruckner and Catalano 2018; Goin et al. 2021; Raz et al. 2018). Spontaneous abortion has been associated with exposure to OGD; women residing in Ecuadorian communities within 5 km downstream of an oil field had greater odds of spontaneous abortions relative to those living at least 30 km upstream of an oil field (San Sebastian et al. 2002). Because we were not able to examine fetal loss in our analysis, we cannot rule out the possible role of live birth bias in our analysis.

This study had limitations. To assign exposure to each pregnancy period, we used data on verified HF well status. Most HF occurs in Kern County, but only 20% of Kern County well records were randomly sampled to verify HF status prior to 2014; we underestimated the number of HF wells and women exposed, likely biasing effect estimates towards the null. We could not fully evaluate the impact of missingness on our results without an accurate probability of HF for births with missing data. While missingness could have biased our effect estimates in any direction, the impact is likely to be minimal as only 5% of study county births in our 8 county study area would have occurred in Kern where stimulation is most likely to occur compared to the 7 other counties. With a limited number of exposed births, we were also unable to assess trimester-specific effects. Additionally, the HF well data did not include specific dates for phases of pre-production (i.e., pad development, drilling, and stimulation) which precluded assessment of hazards at each phase of well creation or stimulation. Another limitation was our reliance on distance to HF wells as a proxy for exposure to diverse HF hazards that have yet to be fully characterized. However, distance allows evaluation of associations for large populations and serves as an aggregate measure for potential physical, chemical and social stressors associated with HF, and can inform regulations such as minimum allowable distances to well sites (Deziel 2021). Finally, we did not have access to data on maternal occupation, BMI, smoking status, or maternal mobility during pregnancy, which likely modestly biased results towards the null (Blanchard et al. 2018; Chen et al. 2010; Hodgson et al. 2015; Lupo et al. 2010; Pennington et al. 2017).

Our retrospective birth cohort study, the first study of HF in California, adds to the evidence that prenatal exposure to HF is associated with adverse birth outcomes. Relative risk is high although absolute risk may be low across the state. While findings from this study may not be generalizable and additional studies are needed to verify these findings, results from this and our previous work can inform regulatory strategies in California and motivate research to better characterize potential HF-specific hazards and the adequacy of current setback distances to OGD, and HF in particular, especially in rural areas.

2.6 Tables

Table 2.1. Neonate, maternal and area-level characteristics of 2006-2015 births by binary hydraulic fracturing (HF) exposure category in eight California counties with HF wells. The percentage is provided unless otherwise indicated in the variable column. Note that active wells include all wells that produced oil or gas during our study period while inactive wells did not produce anything. Only wells within 1 km of residences were counted.

Variable	N (%) 1,005,755	No HF wells (%) n=1,004,563	HF wells (%) n=1,192	p-value†
Neonate characteristics				
Mean birth weight (g) (SD)	3310 (523)	3,310 (523)	3,304 (545)	0.54
Mean gestational age (weeks) (SD)	39.1 (2.0)	39.2 (2.0)	39.3 (1.9)	0.008
Low birth weight	52,378 (5)	5	5	0.90
Preterm birth	70,772 (7)	7	5	0.01
Small for gestational age	120,590 (12)	12	16	<0.0001
Missing	4 (<0.01)	<0.01	0	
Conception year				
2005	81,081(8)	8	8	<0.0001
2006	109,838 (11)	11	21	
2007	108,906 (10)	11	24	
2008	103,191 (10)	10	3	
2009	97,253 (10)	10	1	
2010	96,915 (10)	10	8	
2011	95,498 (9)	9	9	
2012	96,446 (10)	10	12	
2013	97,472 (10)	10	9	
2014	95,526 (10)	9	4	
2015	23,629 (2)	2	1	
Maternal Characteristics (%)				
Education				
< High school	278,658 (28)	28	23	<0.0001
High school diploma/GED	241,528 (24)	24	20	
Some college	221,485(22)	22	24	

College+	238,535 (24)	24	31	
Missing	25,549 (2)	2	2	
Age at delivery				
< 20	84,400 (8)	8	8	0.04
20-24	208,964 (21)	21	18	
25-29	261,529 (26)	26	25	
30-34	259,815 (26)	26	29	
35+	191,042 (19)	19	20	
Missing	5 (<0.01)	<0.01	0	
Race/ethnicity				
Asian/Pacific Islander	128,273 (13)	13	12	<0.0001
Black	43,829 (4)	4	16	
Hispanic	602,738 (60)	60	50	
Other	23,048 (2)	2	4	
White	207,867 (21)	21	18	
Kotelchuck index				
Inadequate	101,192 (10)	10	13	0.0004
Intermediate	97,007 (10)	10	11	
Adequate+	337,530 (33)	33	31	
Adequate	470,026 (47)	47	45	
Parity				
Nulliparous	392,327 (39)	39	42	0.03
Multiparous	612,989 (61)	61	58	
Missing	439 (<0.01)	<0.01	<0.01	
Area-level characteristics (%)				
County				
Colusa	1,755 (0.2)	0.2	0.3	<0.0001
Fresno	131,406 (13)	13	0.3	
Glenn	1,730 (0.2)	0.2	0.2	
Kern	34,305 (3)	3	17	
Los Angeles	573,911 (57)	57	65	
Orange	198,259 (20)	20	7	

Santa Barbara	33,157 (3)	3	0.1	
Ventura	31,232 (3)	3	10	
Mean annual NO_2 (ppb) (SD)	18 (7)	18 (7)	18 (8)	0.37
Missing	2 (<0.01)	<0.01	0	
Urban	936,724 (93)	93	80	<0.0001
ICE for income				
Quartile 1 - poverty	251,667 (25)	25	35	<0.0001
Quartile 2	250,933 (25)	25	12	
Quartile 3	252,021 (25)	25	16	
Quartile 4 - wealth	251,092 (25)	25	37	
Missing	42 (<0.01)	<0.01	0	
Wells				
Mean active+inactive well count (SD)[a,b]	0.2 (7)	0 (0)	143 (148)	
Mean inactive well count (SD)[a]	0.1 (5)	0 (0)	98 (104)	
Mean active well count (SD)[b]	0.1 (2)	0 (0)	45 (51)	
Mean BOE/day of gestation (SD)	1 (66)	0 (0)	1,089 (1,583)	

g grams; *SD* standard deviation; *ppb* parts per billion; *HF* hydraulic fracturing; *ICE* Index of Concentration at the Extremes

BOE barrels of oil equivalent (gas cubic feet converted to BOE to sum to barrels of oil)

†ANOVA or chi-square test

[a]Well count within 1 km of residences across pregnancy and derived by taking the difference between total well count and active well count within 1 km.

[b]Well count within 1 km of residences across pregnancy and based on whether a well had monthly production volume.

Table 2.2. Adjusted odds ratios and mean difference for adverse birth outcomes associated with exposure to hydraulic fracturing (HF) during pregnancy by urban and rural census tract for the primary analysis using a reference group of 2006-2015 births to mothers who were not exposed to any oil or gas wells within 1 km across the eight California counties (N= 979,961) (Model 1).

	No wells (ref)		1+ HF wells			
	n	Cases (%)	*n*	Cases (%)	EE (95% CI)	EM p-value[c]
Low birth weight[a]						
Rural	66,822	3,183 (5)	225	15 (7)	1.74 (1.10, 2.75)	0.007
Urban	911,977	47,761 (5)	937	45 (5)	0.83 (0.63, 1.07)	
Preterm birth[a]						
Rural	66,822	4,903 (7)	225	13 (6)	1.17 (0.64, 2.12)	0.09
Urban	911,977	64,048 (7)	937	48 (5)	0.65 (0.48, 0.87)	
Small for gestational age[a]						
Rural	66,822	7,237 (11)	225	40 (18)	1.68 (1.42, 2.27)	0.10
Urban	911,977	110,146 (12)	937	144 (15)	1.23 (0.98, 1.55)	
Term birth weight (g)[b]						
Rural	61,919	--	212	--	-73 (-131, -15)	0.05
Urban	847,929	--	889	--	-2 (-35, 31)	

EE effect estimate; *CI* confidence interval; *EM* effect modification; *g* grams

Note: Eight counties included: Colusa, Fresno, Glenn, Kern, Los Angeles, Orange, Santa Barbara and Ventura

[a]Logistic regression models (odds ratio) with generalized estimating equations adjusted for child's sex, conception month and birth year;

maternal education, age, race/ethnicity, Kotelchuck prenatal care index, parity; urban indicator, NO_2 concentration, air basin, and ICE for income.

[b]Linear regression model (mean difference) with generalized estimating equations also adjusted for gestational age in addition to those in footnote a.

[c]Test for difference in strata-specific effect estimates between rural and urban populations. Effect modification p-values were derived from

two-sample z-tests using strata-specific estimates and variances.

2.7 Figures

Figure 2.1. Hydraulic fracturing (HF) well density within Kern County (2005–2015), where 90% of HF in California occurred between 2005 and 2015. Seven other counties were included in this analysis but we zoomed into the county with the highest occurrence of hydraulic fracturing. The map was created in ArcGIS 10.6 (Esri, Redlands, CA). Well density was calculated via the point density tool, based on the number of neighboring wells within a 1 km x 1 km cell around each well.

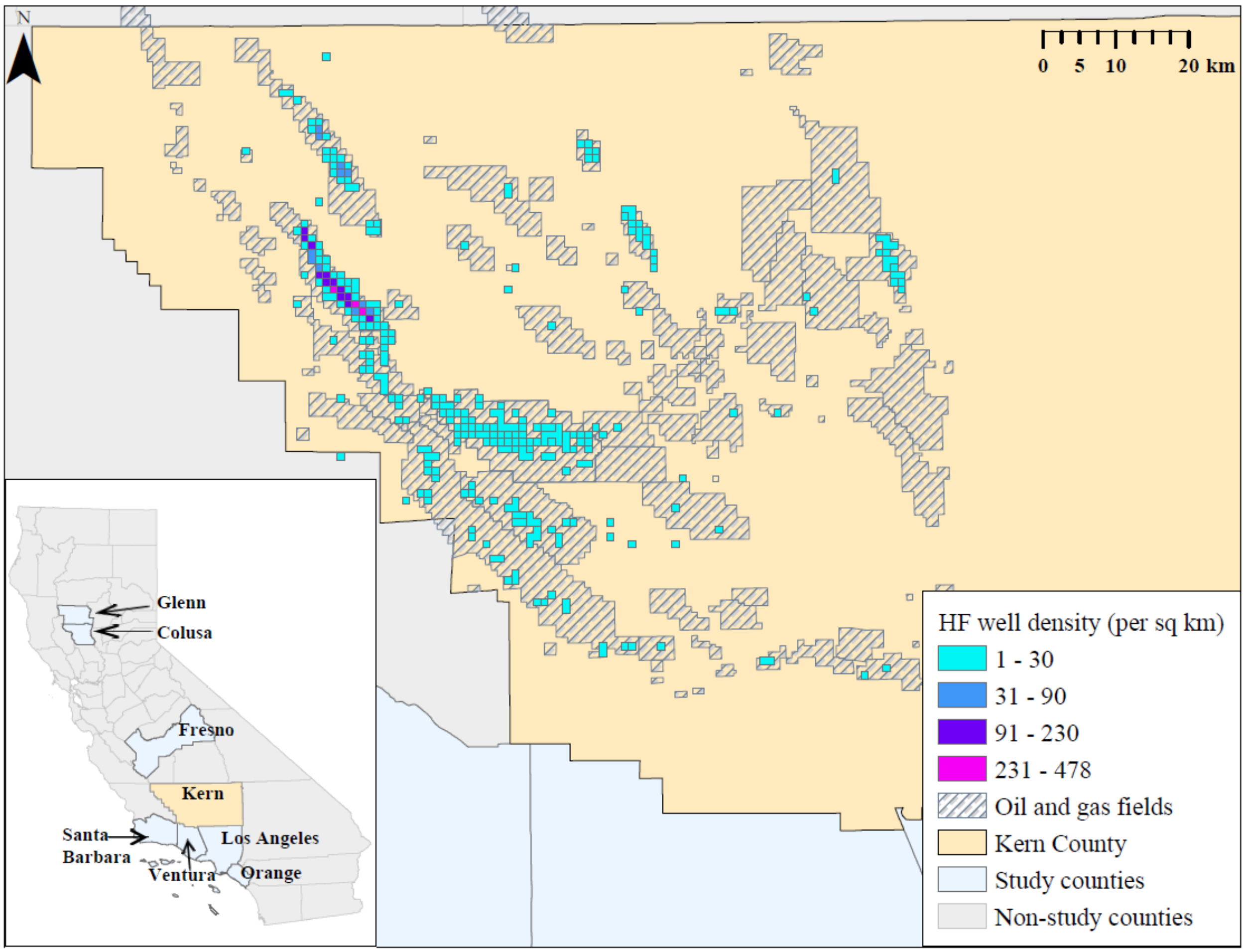
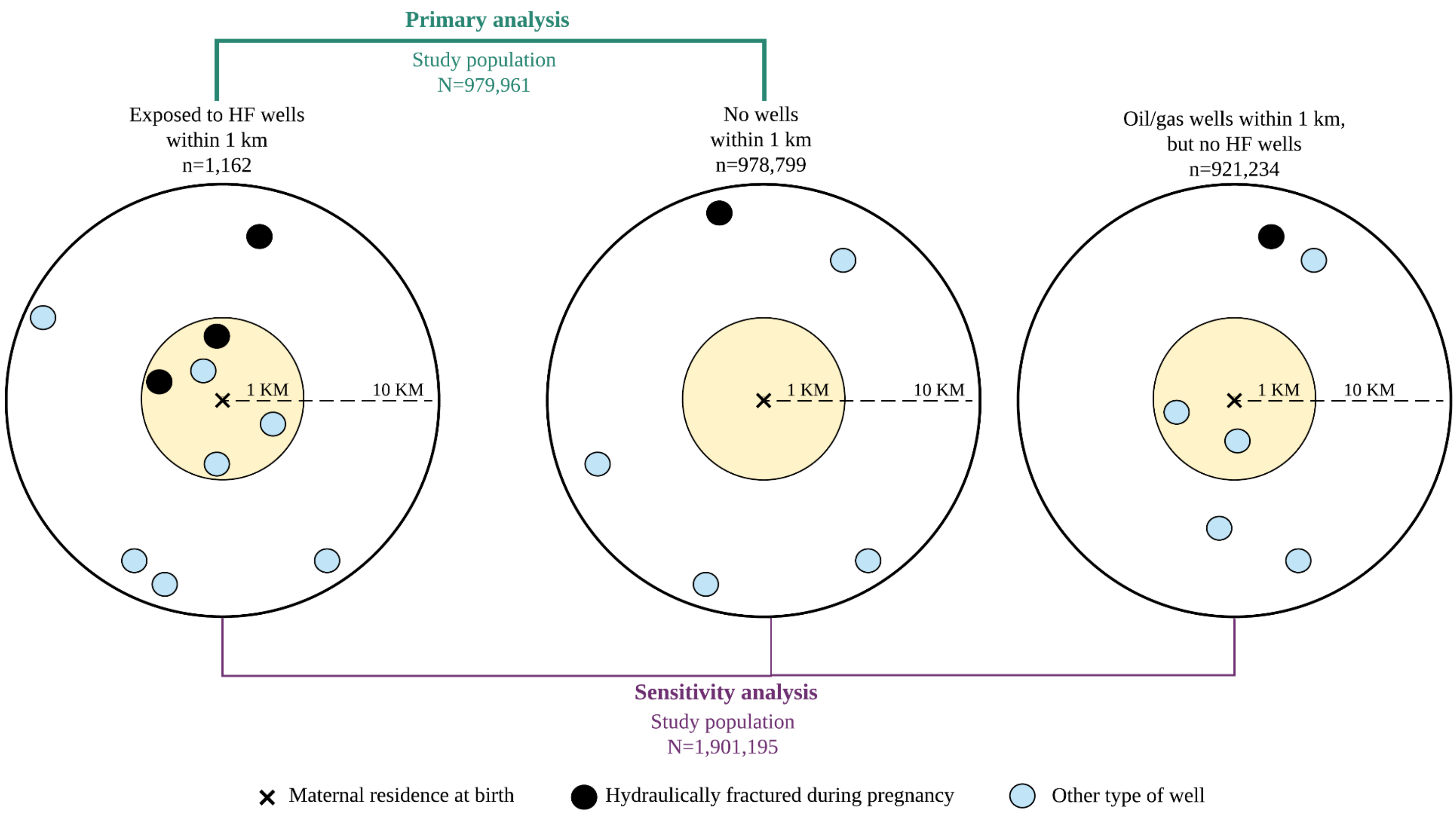
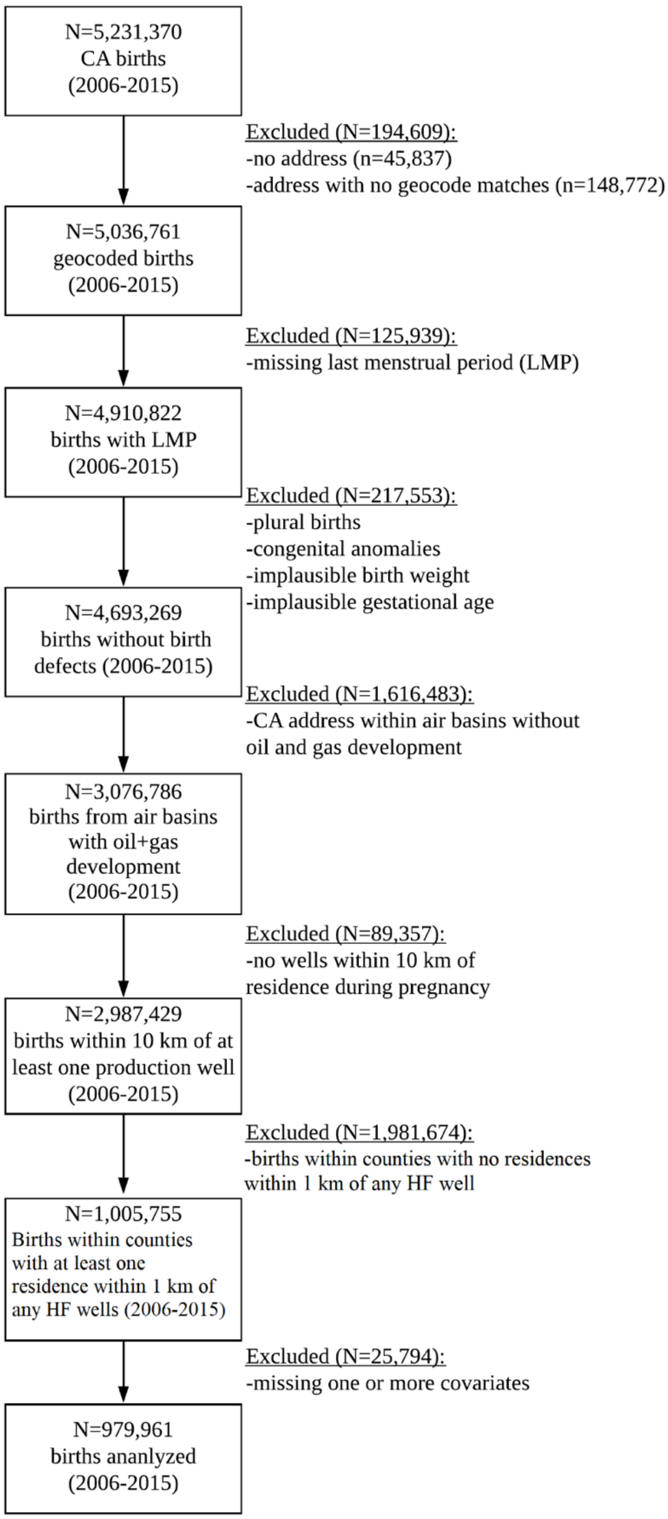


Figure 2.2. Schematic of exposed and reference groups for the primary and sensitivity analyses. For both primary and secondary analyses, exposed mothers had at least one well that was hydraulically fractured during pregnancy within 1 km of maternal residence. For the primary analysis, reference mothers had no oil or gas wells of any kind within 1 km of maternal residence during pregnancy. For the sensitivity analysis, the reference group consisted of mothers without HF within 1 km of maternal residence during pregnancy, including women who lived within 1 km of no wells and women who lived within 1 km of at least one oil or gas well that was not recorded as being hydraulically fractured during their pregnancy.



2.8 Supplemental Information Chapter 2

Supplemental Figure 2.1. Flow diagram of study population development and exclusion criteria applied.



Supplemental Table 2.1. Unstratified adjusted odds ratios and mean difference (grams) for birth outcomes associated with exposure to hydraulically fractured wells during pregnancy for Model 1 with a reference group of 2006-2015 births to mothers without any oil or gas well exposure and Model 2 with a reference group consisting of 2006-2015 births to mothers that were exposed to no wells or wells that were not HF within 1 km across the eight California counties.

	No HF wells (ref)		1+ HF wells		
	n	Cases (%)	*n*	Cases (%)	EE (95% CI)
Model 1					
Low birth weight[a]	978,799	50,944 (5)	1,162	60 (5)	0.95 (0.74, 1.21)
Preterm birth[a]	978,799	68,951 (7)	1,162	61 (5)	0.71 (0.54, 0.93)
Small for gestational age[a]	978,799	117,383 (12)	1,162	184 (16)	1.31 (1.08, 1.59)
Term birth weight (g)[b]	909,848	--	1,101	--	-15 (-46, 16)
Model 2 (sensitivity)					
Low birth weight[c]	1,900,033	97,822 (5)	1,162	60 (5)	0.95 (0.74, 1.21)
Preterm birth[c]	1,900,033	130,564 (7)	1,162	61 (5)	0.72 (0.54, 0.95)
Small for gestational age[c]	1,900,033	226,892 (12)	1,162	184 (16)	1.32 (1.08, 1.61)
Term birth weight (g)[d]	1,769,469	--	1,101	--	-17 (-48, 13)

EE effect estimate; *CI* confidence interval; *g* grams.

Note: Model 1, no wells reference group; Model 2, all 2006-2015 births from 8 counties with HF; Eight counties analyzed: Colusa, Fresno, Glenn, Kern, Los Angeles, Orange, Santa Barbara and Ventura.

[a]Logistic regression models (odds ratio) with generalized estimating equations adjusted for child's sex, conception month and year; maternal education, age, race/ethnicity, Kotelchuck prenatal care index, parity; urban indicator, NO_2 concentration, air basin, and ICE for income.

[b]Linear regression model (mean difference) with generalized estimating equations adjusted for gestational age in addition to those in footnote a.

[c]Logistic regression models (odds ratio) with generalized estimating equations adjusted for active production volume, inactive well count; child's sex, conception month and year; maternal education, age, race/ethnicity, Kotelchuck prenatal care index, parity; urban indicator, NO_2 concentration, air basin, and ICE for income.

[d]Linear regression model (mean difference) with generalized estimating equations adjusted for gestational age in addition to those in footnote c.

Supplemental Table 2.2. Odds ratios and mean difference (grams) for birth outcomes associated with exposure to hydraulically fractured wells during pregnancy stratified by urban/rural status and adjusted for one covariate at a time. The four covariates modeled individually are maternal age, education, race/ethnicity and LMP year.

	No wells (ref)		1+ HF wells		Effect Estimates (95% Confidence Interval)			
	n	Cases (%)	*n*	Cases (%)	Mat Age	Mat Education	Mat Race/ethnicity	LMP Year
Low birth weight[a]								
Rural	66,822	3,183 (5)	225	15 (7)	1.49 (0.99, 2.24)	1.46 (0.98, 2.18)	1.37 (0.87, 2.17)	1.41 (0.92, 2.16)
Urban	911,977	47,761 (5)	937	45 (5)	0.92 (0.69, 1.23)	0.93 (0.70, 1.24)	0.80 (0.61, 1.04)	0.93 (0.70, 1.24)
Preterm birth[a]								
Rural	66,822	4,903 (7)	225	13 (6)	0.86 (0.45, 1.63)	0.81 (0.46, 1.43)	0.82 (0.44, 1.52)	0.89 (0.46, 1.73)
Urban	911,977	64,048 (7)	937	48 (5)	0.72 (0.52, 0.99)	0.74 (0.54, 1.00)	0.66 (0.49, 0.90)	0.68 (0.51, 0.92)
Small for gestational age[a]								
Rural	66,822	7,237 (11)	225	40 (18)	1.87 (1.43, 2.45)	1.83 (1.41, 2.37)	1.66 (1.23, 2.24)	1.80 (1.41, 2.31)
Urban	911,977	110,146 (12)	937	144 (15)	1.33 (1.05, 1.68)	1.33 (1.06, 1.68)	1.21 (0.97, 1.52)	1.33 (1.06, 1.68)
Term birth weight[a]								
Rural	61,919	--	212	--	-88 (-169, -8)	-83 (-162, -4)	-63 (-126, 0.40)	-82 (-158, -6)
Urban	847,929	--	889	--	-13 (-51, 25)	-15 (-53, 23)	4 (-33, 40)	-16 (-56, 22)

Mat maternal; *LMP* last menstrual period

[a]Regression models (odds ratio/mean difference) with generalized estimating equations adjusted for one covariate.

Supplemental Table 2.3. Adjusted odds ratios and mean difference (grams) for birth outcomes associated with exposure to hydraulically fractured wells during pregnancy by urban/rural status for the sensitivity analysis using a reference group consisting of 2006-2015 births to mothers that were exposed to no wells or wells that were not HF within 1 km across the eight California counties. Total production volume (sum of monthly volume) and inactive well count (difference between monthly count of all wells within 1 km and active wells within 1 km) across pregnancy were also adjusted for in these models.

	No HF wells (ref)		1+ HF wells			
	n	Cases (%)	*n*	Cases (%)	EE (95% CI)	EM p-value[c]
Low birth weight[a]						
Rural	118,806	5,438 (5)	225	15 (7)	1.47 (0.86, 2.53)	0.07
Urban	1,781,227	92,384 (5)	937	45 (5)	0.83 (0.64, 1.09)	
Preterm birth[a]						
Rural	118,806	8,177 (7)	225	13 (6)	1.14 (0.65, 1.98)	0.10
Urban	1,781,227	122,387 (7)	937	48 (5)	0.65 (0.47, 0.90)	
Small for gestational age[a]						
Rural	118,806	12,683 (11)	225	40 (18)	1.60 (1.17, 2.18)	0.18
Urban	1,781,227	214,209 (12)	937	144 (15)	1.24 (0.98, 1.57)	
Term birth weight (g)[b]						
Rural	110,629	--	212	--	-49 (-103, 4)	0.22
Urban	1,658,840	--	889	--	-7 (-41, 27)	

EE effect estimate; *CI* confidence interval; *EM* effect modification; *g* grams.
Note: Eight counties included: Colusa, Fresno, Glenn, Kern, Los Angeles, Orange, Santa Barbara and Ventura.
[a]Logistic regression models (odds ratio) with generalized estimating equations adjusted for active production volume, inactive well count; child's sex, conception month and birth year; maternal education, age, race/ethnicity, Kotelchuck prenatal care index, parity; urban indicator, NO_2, air basin, and ICE for income.
[b]Linear regression model (mean difference) with generalized estimating equations also adjusted for gestational age in addition to those in footnote a.
[c]Test for difference in strata-specific effect estimates between rural and urban populations. Effect modification p-values were derived from two-sample z-tests using strata-specific estimates and variances.

Chapter 3: Assessment of potential drinking water threats posed by oil and gas development sites in the San Joaquin Valley, California

3.1 Abstract

Background: California's San Joaquin Valley (SJV) is vulnerable to many environmental disparities and an epicenter of oil and gas development (OGD) in the state. Contamination of limited groundwater resources can occur in the production and disposal phases, which involve the use of production wells, Class II injection wells to facilitate production enhancement and wastewater disposal, and percolation pits for disposal. Additives and naturally occurring chemicals and compounds could leach into waterways via spills, leaks, equipment failures, direct percolation or intended and unintended fractures.
Objective: To first characterize the spatial relationships between OGD infrastructure and domestic wells areas (populated areas served by at least one domestic well–DWA) and community water systems (public drinking water systems with at least 15 connections–CWS) in order to identify potential groundwater threats to DWA and CWS and then determine whether at-risk drinking water sources in the SJV serve vulnerable populations.
Methods: We evaluated DWA and CWS separately. We first identified the number and type of OGD wells and pits within 3 km of DWA and 1 km of CWS. Active OGD well types included those for oil and gas production, hydraulic fracturing, enhanced oil recovery (EOR) and wastewater disposal. Unlined pits include those that were active, inactive, or closing. We then applied regression models to determine whether at-risk drinking water sources served vulnerable populations. DWA were analyzed at the block group level. We modeled the total count of all OGD wells and pits (dependent variable) with several DWA block group or CWS demographic and other area-level factors. We adjusted for block group or CWS area via an offset term. The negative binomial generalized additive model (GAM) and negative binomial hierarchal generalized linear mixed model (GLMM) were used to adjust for spatial autocorrelation among block groups (DWA) and CWS, respectively.
Results: Among 492 block groups with DWA, 61 (12%) had OGD within 3km, while 51 (12%) of 417 CWS's had OGD within 1 km. Block groups that intersected with OGD had fewer residents using domestic wells per km2 (24 vs. 63), a higher proportion of residents living at twice below poverty (50 vs. 45%) and a higher proportion of Hispanics (59 vs. 49%), on average, relative to block groups that did not intersect OGD. Ten DWA block groups with OGD infrastructure also intersected with the Monterey Shale. Block groups with OGD had a mean of 41 total wells and pits, with high counts of actively producing and EOR wells. On average, CWS that intersected OGD infrastructure had fewer residents within the system per km2 (645 vs. 817), a lower proportion of residents living twice below poverty (37 vs. 42%) and higher proportion of Hispanics (47 vs. 39%) compared to CWS that did not intersect OGD. Four CWS with OGD infrastructure also intersected with the Monterey Shale. CWS with OGD had a mean of 21 total wells and pits, with high counts of actively producing and EOR wells as well. No socioeconomic (SES) or area-level factors significantly predicted the count of OGD wells and pits per km2. However, effect estimates indicate differences in the distribution of OGD infrastructure and SES factors across water systems. For CWS, small systems significantly predicted higher counts of OGD infrastructure compared to larger systems [223 (95% CI: 22.1, 2247) times the mean number of OGD wells and pits per km2].

Conclusion: Small CWS and the populations they serve are at greater risk for potential drinking water quality threats due to proximity to OGD infrastructure. Our work highlights the need for increased groundwater source monitoring, especially within small CWS, and standardization of regulatory strategies for protecting drinking water sources from potential contamination (e.g. aquifer exemptions and water sampling requirements) across all OGD infrastructure.

3.2 Background

California's Central Valley is one of the most environmentally impacted regions of the state, with a significant proportion of poor residents and communities of color. The San Joaquin Valley (SJV) makes up the southern portion of the Central Valley and consists of eight counties: San Joaquin, Stanislaus, Madera, Merced, Fresno, Tulare, Kings, and Kern. According to Cal-EPA's CalEnviroScreen, a regulatory tool that identifies vulnerable communities disproportionately burdened by multiple sources of pollution and guides decision-making to address environmental justice, most SJV census tracts have scores within the 60-100th percentile, indicating that these tracts have high pollution burdens and high proportions of socially vulnerable populations (OEHHA 2016).

Poor drinking water quality, antiquated infrastructure, and limited water resources are major environmental and public health issues within the region. Aquifers are particularly stressed by high usage by the agricultural sector. Groundwater is also the primary source of drinking water for about one million SJV residents (Smith et al. 2018). Water contaminants of concern in the region include arsenic, nitrate and salts that cause water sources to become brackish and undrinkable (Balazs Carolina et al. 2011; Balazs et al. 2012a; Hanak et al. 2019). Contaminant sources are geogenic (i.e. naturally occurring) and anthropogenic (e.g. pesticides and fertilizers used in agriculture, oil and gas contaminants) (Hanak et al. 2019). High demands and drought are exacerbating contamination issues, as aquifers are over-pumped, resulting in subsidence and an increase in aquifer arsenic concentrations (Smith et al. 2018). Additionally, socially and economically marginalized communities (e.g. renters, Hispanics) face disproportionate contaminant exposures, such as to arsenic and nitrate, and challenges of unequal compliance with drinking water regulations (Balazs Carolina et al. 2011; Balazs et al. 2012a). Because the SJV is an epicenter for oil and gas development, the industry presents additional threats to drinking water resources; accordingly, we sought to characterize the relationship between groundwater-based drinking water systems and the presence of oil and gas development (OGD) infrastructure in the SJV.

Broadly, OGD consists of production, disposal and storage phases. Here, we focus on infrastructure elements related to production and disposal but not hydrocarbon storage processes. The production phase involves the use of wells that extract oil and gas from sub-surface levels and injection wells to enhance production. Injection of steam, water, air, brine (or wastewater), polymers and carbon dioxide are traditional enhanced oil recovery (EOR) methods that change the sub-surface pressure and hydrocarbon properties, e.g. viscosity, to improve permeability and direct the hydrocarbons toward production wells (Millemann et al. 1982; US EPA 2015b). Well stimulation with chemicals (i.e. hydraulic fracturing (HF), acid fracturing and matrix acidization) is an unconventional enhancement technique that requires the injection of fluids mixed with chemicals at high pressure to induce fractures within geological formations to release the oil and gas (US EPA 2015b). HF is the most common unconventional practice. In California, most oil and gas production is facilitated by some form of EOR (Long et al. 2015a). The disposal phase involves transferring wastewater composed of produced water and flowback via pipes or trucks to disposal injection wells and percolation pits (or ponds, sumps, impoundments) (Long et al. 2015b; US EPA 2016, 2015). US EPA classifies all oil and gas injection wells for production and disposal purposes as Class II injection wells. Produced water, or formation water, consists of naturally occurring compounds from the geologic formation and arises along with oil and gas in traditional and unconventional production operations. Flowback is the proportion of water injected during well stimulation operations that returns to the surface prior to oil and gas production. Wastewater can be composed of: 1) dissolved substances from formation water, 2) substances mobilized from the target geological formation, 3) some residual oil and gas, 4) additives pumped into the well during well stimulation and 5) compounds that formed due to chemical reactions between additives or due to transformation or degradation of the additives (Long et al. 2015a; US EPA 2016). Contamination of drinking water sources with these formation and wastewater constituents may harm human health. Although biological mechanisms and specific pollutants have not yet been determined, epidemiology studies have linked exposure to OGD with various adverse health outcomes, particularly birth outcomes (Deziel et al. 2020; Gonzalez et al. 2020a; Tran et al. 2020a).

Contaminants can enter drinking water sources through several pathways: 1) surface spills of wastewater, HF fluid or raw additives from equipment failure or human error; 2) leaks from all types of production and injection wells due to poor construction or natural degradation of well casings over time; 3) broken pipelines that transfer produced water for disposal; 4) percolation of fluid in pits with compromised lining or that are unlined; 5) fractures created during stimulation operations that were directly start in or grow into overlying aquifers (increased likelihood with shallow wells); and 6) pathways that stray out of the oil/gas reservoir (out-of-zone) due to stimulation or EOR operations that connect with a preexisting fracture network, a fault, or some other permeable feature (Long et al. 2015b; McIntosh and Ferguson 2019; Millemann et al. 1982; US EPA 2016). While EOR does not involve a large number of chemicals, impacts can occur over a long period. Well stimulation operations occur within several hours to days (McIntosh and Ferguson 2019). In contrast, traditional EOR wells operate under low pressure for much longer time periods. For example, the median operation period for EOR wells in Western Canada was 35 years (McIntosh and Ferguson 2019). Older wells can pose greater threats to drinking water sources as chances of well failures increase with age. There is also a risk that the sub-surface pressure created by EOR operations can reach a point where unintended fractures develop and expand into water sources over time (McIntosh and Ferguson 2019; Millemann et al. 1982).

Spills and well failures are regularly reported to local and state level regulatory agencies, but evidence of impact to water sources has focused on unconventional OGD. Impacts to drinking water resources are determined by the characteristics of spills/leaks/fractures, local geology, and the fate, transport, and toxicity of chemicals spilled (US EPA 2016). Wastewater typically consists of HF additives, salts, trace elements (e.g. strontium, barium, heavy metals), organics (e.g. benzene, toluene, ethylbenzene and xylene (BTEX)), and NORMs or naturally occurring radioactive materials (e.g. radium-226) (Long et al. 2015b). Organic compounds such as BTEX, alcohols and chloride compounds as well as heavy metals have been found in groundwater samples collected near primarily unconventional OGD sites in Texas, Pennsylvania, Colorado,

Louisiana and Arkansas (Burton et al. 2016; Drollette et al. 2015; Gross et al. 2013b; Hildenbrand et al. 2015, 2016; Llewellyn et al. 2015; McMahon et al. 2017; US EPA 2016). Across studies, most concentrations were low but some exceeded maximum contaminant levels (MCL). Researchers attributed these contaminants to projected or recorded surface spills (Drollette et al. 2015; Gross et al. 2013b; Hildenbrand et al. 2016; McMahon et al. 2017), compromised cement casing (Burton et al. 2016), or out-of-zone vertical fractures (Llewellyn et al. 2015). In the Colorado study where spill events and groundwater samples were analyzed, authors also found that remediation efforts were able to reduce impacts (Gross et al. 2013b). Sub-surface impacts are more challenging to track and assess due to lower flow rates, decreased mixing and potential delays in detecting leaks from the wells and out-of-zone fractures.
In addition to well failures, injection wells may be near groundwater aquifers that should be protected as an "underground source of drinking water" (USDW) as stipulated by the Safe Drinking Water Act (SDWA). The US Environmental Protection Agency (EPA) defines USDWs as containing less than 10,000 mg/L total dissolved solids (TDS), i.e. salts. Aquifers with high TDS (more than or equal to 10,000 mg/L) and that are not likely to serve as a source of drinking water can receive an exemption from being protected by the US EPA (US EPA 2015c). CalGEM exempts those injection wells overlying exempt aquifers based on the US EPA's USDW criteria (DiGiulio and Shonkoff 2019; Long et al. 2015a). In 2011, a US EPA audit identified many injection wells in California that were located over aquifers that had yet to be evaluated for exemption by the Agency. Reasons for these violations are two-fold: first, aquifer TDS had not been assessed for the wells already permitted and second, the aquifer exemption TDS criteria were different in California. Prior to the audit, CalGEM protected aquifers containing less than 3,000 mg/L TDS so water sources with TDS between 3,000 and 10,000 mg/L were not being protected as potential sources of drinking water according to the EPA's criteria (Long et al. 2015b). CalGEM has since adopted EPA's exemption criteria for all Class II injection wells (including stimulation operations), updated permitting procedures, and reviewed injection wells identified by the EPA's audit. During this review process, water samples collected from wells within a mile of several non-exempt injection wells did not indicate local water contamination by oil and gas (Long et al. 2015b). However, there could be historic contaminants that transported beyond one mile or at other well sites that were not tested. The US Geological Survey (USGS) has begun to conduct water quality studies at sites near oil and gas fields to determine where and to what degree groundwater quality is potentially at risk for OGD after the adoption of Senate Bill 4 (SB 4) (Pavley 2013; SWRCB 2015b).

Percolation pits provide a direct pathway for OGD chemicals and compounds to transport into drinking water sources. In California, most pits are located within the SJV and unlined (93% of SJV pits in 2019). While pits are slowly being phased out, those overlying groundwater aquifers with existing and future beneficial uses have been allowed if the wastewater met certain salinity, chloride, and boron thresholds (Long et al. 2015b). In terms of salinity, aquifers with 3,000 mg/L of TDS were protected since 2006 (DiGiulio and Shonkoff 2019). Prior to the passage of SB 4, the Central Valley Regional Water Quality Control Board (CVRWQCB) did not require chemical analysis of wastewater disposed in active pits within the Central Valley. The testing requirement, enacted in April 2015, focuses on TDS, chloride and boron but does not include any other wastewater constituents (Long et al. 2015b). Groundwater contamination from percolation pits has been documented in California and other states (Long et al. 2015a; US EPA 2016). Most recently, CVRWQCB shut down a wastewater disposal center that disposed 2.8 million gallons of produced water per day via 163 acres of unlined disposal pits. The Board found that a highly saline wastewater plume from decades of disposing produced water migrated to higher-quality groundwater for municipal and agricultural uses (CVRWQCB 2019b). The extent of groundwater impacts from unlined pits is currently unknown but CVRWQCB have increased efforts to monitor active pits.

Several studies have sought to characterize and identify potentially vulnerable communities within proximity to OGD via spatial methods. Methods range from mapping and identifying clusters based on co-location of OGD infrastructure and indicators for vulnerable populations (e.g. people of color, poverty) via local indicators of spatial association (LISA) analysis (Ogneva-Himmelberger and Huang 2015a) or clusters of intense OGD and population at risk via developing an intensity function (Meng 2015b); comparing differences in socioeconomic factors of populations within proximity to OGD wells and beyond via t-tests (Clough and Bell 2016a); and determining socioeconomic predictors of areas with OGD sites via fixed effect modeling (Zwickl 2019), Poisson regression (Johnston et al. 2016a), and spatial regression techniques including conditional autogressive (Silva et al. 2018) and generalized additive modeling (Johnston et al. 2020a). The units of analysis in these studies vary from wells, census blocks, block groups and tracts, with buffers of up to 5 km. Studies focused on OGD in Texas (Johnston et al. 2016a, 2020a; Zwickl 2019), Pennsylvania (Clough and Bell 2016a; Meng 2015b; Ogneva-Himmelberger and Huang 2015a; Zwickl 2019), Ohio (Ogneva-Himmelberger and Huang 2015a; Silva et al. 2018), West Virginia (Ogneva-Himmelberger and Huang 2015a), Oklahoma and Colorado (Zwickl 2019). No studies have assessed socioeconomic disparities in California or spatial relationships of OGD relative to drinking water systems. Meng et al (2015) buffered wells rather than water systems in order to assign risk to wells, but did not examine the characteristics of population using those systems.

Groundwater feeds into all types drinking water systems but is the only source for private domestic wells and the primary source for some community water systems (CWS). Private domestic wells might serve one person or household up to less than 15 service connections (SWRCB 2015a). CWS are defined as drinking water systems with at least 15 service connections used by yearlong residents or that regularly serve at least 25 yearlong residents of the area served by the system (SWRCB 2012). Domestic wells are not regulated and not regularly monitored, as agencies only provide guidance on contaminants to test for and water quality treatment options (SWRCB 2015a). As such, domestic well users are highly vulnerable to potential contamination since water quality can vary greatly from public systems that are required to meet regulatory standards. Additionally, per SB 4 requirements, domestic well owners near oil and gas fields with stimulation operations can request water sampling but annual CalGEM reports have documented few such requests (CalGEM 2015). CWS are required to have adequate technical, managerial and financial (TMF) capacity to meet drinking water standards established by the SDWA (Soelter and Miller 1999). However, TMF capacity varies greatly by the size of CWS. Small systems (15-199 service connections) often suffer from low TMF capacity as they are operated with limited resources and managed by residents who are unfamiliar with drinking water requirements and planning to ensure standards are met. While state funding is available to improve CWS, small systems often do not meet the requirements that require them to already having sufficient TMF capacity and be "shovel ready" because they already lack resources (Balazs and Ray 2014). Smaller CWS are thus more vulnerable to groundwater threats compared to larger systems with greater TMF capacity to handle contamination problems.

The goal of our SJV study was twofold: first to characterize spatial relationships between presence of OGD infrastructure and each type of drinking water source (domestic well and CWS) and identify potentially vulnerable sources; and second, determine whether at-risk drinking water sources serve socially vulnerable populations. While previous spatial analyses described spatial relationships of or identified SES factors that distinguished or predicted areas with various types of OGD wells, few have examined OGD threats directly on water systems. We evaluated each type of drinking water source independently.

3.3 Methods

We first identified potential threats to drinking water sources (CWS and DWA) from OGD infrastructure in the SJV and then applied regression models to identify potential demographic and other area-level predictors. Our analysis focused on systems that draw from groundwater only i.e. domestic wells and CWS with groundwater as primary sources. For CWS we excluded purchased water because the source of the water is less likely local. All shapefiles were projected at NAD 1983 California Teale Albers (meters) and spatial variables processed with ArcGIS version 10.8 (ESRI).

Water Geographic Layers

Geographic boundaries with demographic data were obtained from the Water Equity Science Shop (WESS) (Goddard 2019; Pace et al. 2020a, 2020b). WESS is a community-academic partnership between UC Berkeley, San Francisco State University (SFSU), Cal EPA's Office of Environmental Health Hazard Assessment (OEHHA) and the Community Water Center (CWC). The methods are briefly summarized.

Likely domestic well areas (DWA): The goal of deriving this layer was to identify communities in California who are likely to be dependent upon domestic wells for their source of drinking water. The location of most domestic wells within CA correspond to the center of Public Land Surveying System (PLSS) sections, which are approximately 1x1 mile grid squares. The PLSS sections were processed in February 2019 along with parcel data, 2010 census population data, well completion reports, and CWS boundaries to ensure that the domestic well areas were populated, had at least one domestic well and did not overlap with CWS boundaries. Data sources included the Bureau of Land Management, US Census, Department of Water Resources (OSWCR), and Tracking California Water System Service Areas tool. The resulting likely domestic well areas were then aggregated into 2010 census block group boundaries. Population size was aerially apportioned to DWA sections that intersected with more than one census block group. The demographic characteristics of domestic well areas were maintained at the census block group level to minimize inaccuracy and uncertainty (Pace et al. 2020b). The DWA shapefile had block group boundaries and features including the number of domestic wells, population served by DWA's, and SES factors in percentages.

Community water systems (CWS): CWS boundaries were obtained from the Tracking California Water System Service Areas tool and connected to information about the systems including the primary water source (e.g. groundwater, surface water or purchased water) and number of service connections from the State Drinking Water Information System (SDWIS) in January 2019. System size was assigned based on the number of service connections as follows: small (15-199 connections or serving at least 25 people year-round), intermediate (200-3,299 connections), medium (3,300-9,999 connections) and large (10,000+ connections) (Pace et al. 2020a). As CWS intersect multiple block group boundaries, population size was aerially apportioned from 2010 census block data for greater accuracy. Since the highest resolution for SES data from the American Community Survey (ACS) 2012-2016 5-year estimates were available at the block group level, the block-based population size was summed up to the block group and used to weight the SES factors (Goddard 2019). The CWS shapefile had boundaries and features including the number of service connections, population served by CWS's (summed from all blocks intersecting the CWS), and SES factors in percentages.

Oil and Gas Development Infrastructure Layers

OGD infrastructure consisted of four categories of wells and two categories of percolation pits. The well shapefile, "All Wells," was downloaded from the CA Department of Conservation's Geologic Energy Management Division (CalGEM) in November 2019 (CalGEM 2019a). Only wells with "active" status as of November 2019 were included in the analysis. Wells were categorized as oil and gas production wells (OGP), enhanced oil recovery wells (EOR) and waste disposal wells (WD). OGP wells included those coded as dry gas and oil and gas. EOR wells included those coded as air injector, injection, pressure maintenance, cyclic steam, steam flood and water flood. WD wells included those coded as gas disposal and water disposal.

A fourth well category was shallow HF wells with vertical depths less than 600 meters. HF wells become production and injection wells after hydraulic fracturing operations occur. To identify wells that were ever fracked between 2008-2019, HF status, approximate date of fracking and well depth were collected from 1) California Council on Science and Technology's (CCST) well stimulation report (Vol 1, Appendix M) (CCST 2015) and 2) CalGEM's well stimulation treatment (i.e. HF, acid fracturing and matrix acidization) disclosure database (CalGEM 2019b). The CCST dataset provided HF status for 2008-2013 while CalGEM's provided HF status for 2014-2019. The vertical well depth was used to select shallow wells with depth less than 600 m. As of 2015, three quarters of all hydraulic fracturing operations in CA occurred in shallow wells less than 600 m and most of these wells were in SJV (Long et al. 2015b). As protected aquifers exist above HF wells, there is an inherent risk of hydraulic fractures connecting to drinking water aquifers and contaminating them or providing a pathway for water to enter the oil reservoir. Thus, only shallow HF wells were included in our analysis.

Data on percolation pits were obtained from the CVRWQCB. CVRWQCB's Wastewater Disposal Pond List (CVRWQCB 2019a) included latitude, longitude, activity status and lining status. We selected only unlined (93% of all pits) active and inactive or closing pits as of November 2019 for this analysis. Lined pits present less risk for groundwater contamination as the lining is intended to prevent wastewater from percolating into the ground. Contamination could occur if lining becomes faulty. Contrastingly, unlined pits allow wastewater to seep into the ground, carrying chemicals and compounds that may have reacted with one another along with the water. Chemical testing of wastewater disposed in the pits is only required for active wells even though water may still remain in inactive and closing pits (Long et al. 2015b). Thus, inactive and closing pits were considered as potential threats to groundwater here as well.

Information on aquifer exemption was collected from the US Environmental Protection Agency (EPA) (US EPA 2017) and exemption status of EOR and WD wells were retrieved from CalGEM (CalGEM 2019c). While some wells were physically located over exempt aquifers on the map, they were not specified as being exempt within the respective CalGEM datasets. To address this discrepancy, their status was modified based on whether a well point intersected with the exempt aquifer boundary. The same approach was applied to HF wells since their exemption status was not available. Production wells were not included as the current exemption criteria do not apply to conventional OGD wells. We mapped the wells with the exemption status to identify potential regions that need more monitoring or permits reviewed.

Response variable
Potential threats from OGD infrastructure were characterized by quantifying the number of each type of well and pit within a 3 km buffer for DWA and 1 km buffer for CWS. One prior study found that most HF wells (60% of HF wells across 14 states) were within 3 km of at least one domestic well and other studies suggested that OGD chemicals might travel horizontally between 1 and 3 km (Jasechko and Perrone 2017). A smaller buffer was selected for CWS because some are very large, so their likelihood of intersecting with OGD wells and pits is greater. Counts of wells and pits that fell within the buffers were then generated. Because the number of DWA and CWS that intersected with each well or pit type was relatively low, the count of each well and pit type were summed to create a total OGD infrastructure count. This measure was used as the response variable in regression models.

Model predictors
Regression model predictors included percent of the population twice under poverty, percent Hispanic, population density and presence/absence of shale play. Among SES variables available from the DWA and CWS shapefiles, poverty and Hispanics were selected because many SES variables were correlated. Impoverished people and Hispanics have been shown to be vulnerable populations for many environmental threats including water contamination (Balazs et al. 2011, 2012b). Population density, number of people per square kilometer, was calculated based on the population served by DWA or CWS and the area of block groups (for DWA) or CWS. California shale plays include the Monterey and Monterey-Tremblor Shales and the shapefile was retrieved from the US Energy Information Agency (US EIA 2019). Shale play is an area of sedimentary rock containing oil and natural gas, and was included in models as a categorical variable for presence or absence.

Statistical analysis
To identify potential predictors of water systems threated by OGD infrastructure, separate regression models were developed for DWA and CWS. Block groups were the unit of analysis for DWA, the unit for CWS was the water system. For both analyses, the population of block groups and CWS were first limited to SJV counties with any OGD activity. Block groups were further filtered to only include those with at least one DWA section within the block group so that the block groups were more comparable; block groups without any DWA would not have domestic wells at risk of contamination from OGD. For the CWS analysis, we included only those water systems that had groundwater that was not purchased as their primary source. Additionally, block groups and CWS with extreme counts of OGD wells and pits greater than 500 were removed in order to achieve model convergence.

We developed generalized linear models (GLM) and models that adjusted for spatial autocorrelation to assess DWA block groups. First, Poisson and negative binomial models were evaluated for over-dispersion via the Pearson dispersion estimate and model fit based on the Akaike information criterion (AIC) and Moran's I (**Supplemental Table 3.1**). While negative binomial models adjusted for over-dispersion, Moran's I indicated the observations were spatially correlated for DWA block groups. Block groups share boundaries and SES factors of neighboring blocks are more similar to each other than block groups further away, leading to spatial autocorrelation. Thus, spatial regression was performed through the generalized additive model (GAM) for DWA block groups. All models were specified with the total count of all OGD infrastructure as the response variable and percent twice below poverty, percent Hispanic and binary shale play as predictors. Population density was also adjusted for in models and area (km^2) of the block group or CWS was specified as the offset. For all CWS models, system size was added as a binary predictor, with intermediate, medium and large systems merged into one group since the number of systems with OGD were minimal, and population density was log transformed to achieve model convergence.

DWA generalized additive model: To adjust for the spatial autocorrelation between block groups, we used the negative binomial GAM to fit a 2-dimensional thin plate splines smoother on the block group centroid coordinates (latitude and longitude). This smoothing technique to adjust for spatial confounding has been applied in a similar OGD analysis (Johnston et al. 2020a) and air pollution studies (Briggs et al. 2008; Brochu et al. 2011; Padilla et al. 2014; Su et al. 2010). The predictors were all modeled as parametric terms. DWA GLM and GAM modeling were conducted in R version 4.0.0 using the MASS package for GLM modeling and mgcv package for GAM modeling. The Pearson dispersion estimate, AIC and Moran's I were compared across all models (**Supplemental Table 3.1**).

CWS generalized linear model: As with the DWA models, the Pearson dispersion estimate, AIC and Moran's I were compared across all models (**Supplemental Table 3.1**). While the Moran's I test did not indicate spatial autocorrelation for the Poisson model, the AIC and dispersion parameter greatly improved for the negative binomial model compared to the Poisson model so results from the negative binomial model are reported. The GLM modeling was conducted in SAS version 9.4 with Proc GLIMMIX. The Laplace estimation method and Newton-Raphson with ridging optimization technique were specified to achieve model convergence.

3.4 Results

The SJV consisted of 760 block groups and 648 CWS's. Four hundred ninety-two block groups had DWA's within their bounds and 417 CWS's used groundwater as their primary source. OGD infrastructure was distributed along the western region of the study area with a heavy concentration of wells and pits in the southwestern area (**Figure 3.1-3.2**). The northern region had mostly OGP wells. There were between 1 and 547 production and HF wells per km^2 (**Supplemental Figures 3.1-3.4**), 1 and 1,650 injection wells per km^2 (**Supplemental Figures 3.2-3.5**), and 1 and 33 percolation pits per km^2 (**Supplemental Figures 3.3, 3.6**) within SJV. Among active EOR wells and WD wells, 11% and 38%, respectively, were specified as exempt or drilled into exempt aquifers (**Supplemental Figure 3.7**). Most OGD wells were not designated as exempt or coincided with aquifers boundaries that were specified to be exempt areas (**Supplemental Figure 3.7**). One percent of shallow HF wells, 38% of active WD wells and 11% of active EOR wells were specified as exempt. A number of wells and pits that were not specified as exempt were co-located with exempt wells.

DWA assessment

Among the 492 block groups with DWA's, 61 had OGD wells and pits within 3 km of their DWA's (12% of block groups) but three of those block groups were removed from the analysis due to their outlying OGD infrastructure counts (**Table 3.1**). On average, domestic wells spread across more area (109 km^2 vs. 42 km^2) and these wells served more people (685 vs. 541 people) within block groups that intersected OGD infrastructure compared to block groups that did not. The mean percent of DWA within a block group that intersected with OGD was 26%, which served an average of 137 residents using domestic wells. Block groups with OGD had fewer residents using domestic wells per km^2 (24 vs. 63), a higher proportion of residents living twice below poverty (50 vs. 45%) and higher proportion of Hispanics (59 vs. 49%), on average, relative to block groups without OGD. The proportion of block groups that intersected with the Monterey shale was greater among block groups with OGD (17 vs. 2%) and the actual number of block groups was slightly higher relative to block groups without OGD (10 vs. 9). Block groups that intersected with OGD had a mean of 41 total oil and gas wells and pits, 28 OGP wells, 2.5 shallow HF wells, 40 EOR wells, 4 WD wells, 10 active pits and 7 inactive pits (**Supplemental Table 3.2**). Among block groups with OGD, production wells were located within the 3 km of DWA's for all block groups and disposal wells were found within 3 km of DWA's for 52% of block groups. The proportion of block groups that intersected other well and pit types was low. The negative binomial GAM model effectively adjusted for both over-dispersion (dispersion estimate: 0.089) and spatial autocorrelation (Moran's I p-value: 1) and the AIC value indicated a better fit than the GLM Poisson and negative binomial models (**Supplemental Table 3.1**). **Table 3.2** shows the results from GLM and GAM negative binomial models. None of the demographic or area-level factors showed significant associations with the number of OGD wells and pits per km^2 among block groups with DWA. However, the directionality of effect estimates changed for percent Hispanic and shale play between the GLM and GAM models. In the GAM model, every 10% increase in Hispanics decreased the average number of OGD wells and pits per km^2 by 0.91 times. The presence of shale play compared to absence of shale play increased the average number of OGD wells and pits per km^2 by 6.41 times. Percent living at twice below poverty did not affect the mean number of OGD wells and pits per km^2.

CWS assessment

Among the 417 CWS, 51 had OGD wells and pits within 1 km (12% of CWS) but five of those were removed from the analysis due to their outlying OGD well and pit counts (**Table 3.3**). On average, CWS that intersected OGD infrastructure had a greater number of service connections (1,256 vs. 649) and served more people (4,556 vs. 2,006) compared to CWS without OGD. CWS with OGD had fewer residents per km^2 (645 vs. 817), a lower proportion of residents living twice below poverty (37 vs. 42%) and higher proportion of Hispanics (47 vs. 39%) compared to CWS that did not intersect OGD. While the proportion of CWS that intersected with the Monterey shale was greater among CWS with OGD (11 vs. 3%), the actual number of CWS was much lower compared to CWS without OGD (4 vs. 10). Most CWS were small whether they intersected OGD wells and pits or not. CWS that intersected with OGD had a mean of 34 total wells and pits, 21 OGP wells, 87 EOR wells, 3 WD wells, 3.25 inactive pits as of 2019 (**Supplemental Table 3.2**). Among CWS with OGD, production wells were located within 1 km of all CWS and disposal wells were found within 1 km of 61% of CWS. The proportion of CWS that intersected other well and pit types was low.

The negative binomial GLM model effectively adjusted for over-dispersion (dispersion estimate: 0.95) and the AIC value demonstrated a better fit than the GLM Poisson model (**Supplemental Table 3.1**). None of the demographic and area-level factors significantly predicted the number of OGD wells and pits per km^2 among CWS, except system size (**Table 3.4**). Every 10% increase in the proportion of residents living at twice below poverty decreased the mean number of OGD wells and pits per km^2 by 0.64 times. Every 10% increase in Hispanics increased the average number of OGD wells and pits per km^2 by 1.37 times, and presence of shale play also increased the outcome by 1.21 times. Compared to larger CWS, small systems had 223 times (95% confidence interval: 22.1, 2247) the mean number of OGD wells and pits per km^2.

3.5 Discussion

The San Joaquin Valley is a vulnerable region with many threats to water resources, including OGD as the Monterey Shale runs through the entire western zone. While our study had a similar objective of identifying potentially disproportionate impacts of OGD, ours is the first to characterize OGD risks on and SES factors of water systems rather than a census boundary or wells. Our analysis revealed that the relationship between SES factors and number of OGD wells and pits per km^2 varies by the type of water system, likely reflecting the differences in the populations served by each system. We also demonstrated the need to apply different regression techniques depending on the water system.

The direction of associations between predictors and the number of OGD infrastructure contrasted between DWA and CWS except for the shale play variable. Poverty did not predict the number of OGD wells and pits for block groups with DWA, but was associated with a decreased number of OGD infrastructure for CWS. From our maps, CWS with high counts of wells and pits are close to Bakersfield, one of the main urban centers of SJV and near a major oil field. Cities have a mix of high and low poverty census tracts, and the CWS with high OGD well and pit counts are located in areas with lower poverty. Previous oil and gas-related spatial analyses have shown mixed effects: one analysis that buffered unconventional oil and gas wells in Pennsylvania showed poverty was lower within 1 and 2.5 km buffer zones around wells compared to beyond (Clough and Bell 2016a) while another study determined that wastewater disposal wells within 5 km of the census block centroid were sited in high poverty areas of southern Texas (Johnston et al. 2016a). The mean proportion of Hispanics is greater among both DWA block groups with OGD and CWS with OGD compared to those with without, so the positive association aligned with our expectations for CWS. However, increases in Hispanics was associated with lower counts of OGD wells and pits. Based on the maps, most DWA that intersected with high counts of OGD were further away from the city center, suggesting that fewer Hispanics reside in rural areas. Disparities among Hispanics have been observed for the placement of WD wells in Texas (Johnston et al. 2016a), the occurrence of flaring in Texas within 5 km of the census block centroid (Johnston et al. 2020a), and proximity to nearest HF wells in Texas and Colorado (Zwickl 2019). However, the Pennsylvania study also determined that the proportion of Hispanics was lower within the 1 and 2.5 km buffer zones of unconventional wells compared to beyond (Clough and Bell 2016a), which may reflect the overall low proportion of Hispanics in the state. Being located on the Monterey Shale predicted increases in OGD wells and pits among both DWA block groups and CWS with OGD sites. These results were expected since more OGD is likely to occur in areas with shale and a similar effect was observed for shale in the Ohio-based study on WD wells (Silva et al. 2018).

One additional predictor added to CWS models was system size. Our results indicated that small systems were significantly more impacted by OGD than larger systems. Given that medium or large systems have greater area coverage, we would have expected the larger systems to intersect with a higher number of OGD well and pits. These associations suggest that small water systems are more likely to be located near oil and gas fields and at greater risk for potential contamination from OGD infrastructure in addition to having low TMF. The populations small systems serve are, thus, more vulnerable to potential drinking water quality threats. Systems with low TMF often do not meet drinking water quality standards as they lack the resources to consistently test and treat their water. As result of their low TMF, they are usually ineligible to apply for state funding even though these funds are meant to assist high priority projects (Balazs and Ray 2014). This disadvantageous cycle would likely prevent smaller system operators from identifying chemicals from OGD within their system if contamination occurs and remediating the impacts. Smaller systems also serve fewer customers so customers would have to bear a greater burden to pay for water quality tests, treatment and remediation unless operators receive assistance from the state.

The aquifer exemption criteria vary by type of OGD infrastructure and it is notable that boundaries for DWA and CWS intersect aquifers that have been designated by the US EPA as exempt, especially CWS. Furthermore, non-exempt wells and pits are also being operated in areas with domestic wells and CWS. Unconventional and Class II injection wells are subject to the 10,000 mg/L TDS criteria for protected aquifers while percolation pits are subject to the 3,000 mg/L (fresh groundwater) criteria; there are currently no criteria for conventional OGD wells (DiGiulio and Shonkoff 2019). Among water systems that have boundaries over exempt aquifers, wells and pits could potentially become highly salinized at minimum or contaminated with additives and wastewater compounds at worse. Currently exempt and non-exempt wells and pits are co-located, suggesting that not all permitted wells and pits have undergone thorough review since there should be designated exemption zones. Freshwater is defined as water with less than 3,000 mg/L TDS. Water resources with TDS between 3,000 and 10,000 mg/L (brackish water) could have beneficial uses such as drinking water if desalinated, but pits have been permitted to operate over aquifers with TDS within this range (Shonkoff and DiGiulio 2019). Some USGS studies indicate that groundwater has been contaminated to varying degrees, especially in regions of SJV with unlined pits (McMahon et al. 2019; Wright et al. 2019). Increased water sampling and monitoring of domestic wells, especially within small CWS, would help determine whether water is usable from domestic wells in areas with a mix of exempt and non-exempt wells and pits. Water resources with more than 10,000 mg/L TDS could also be useable because of desalination technology (Shonkoff and DiGiulio 2019). However, these aquifers are currently not being protected or tested for OGD contaminants even though drought, climate change and population growth will exhaust current systems.

Domestic and OGD well depth and local TDS levels are important factors we were unable to evaluate in this analysis of drinking water systems that intersect with OGD. The shallower an aquifer, the more likely groundwater sources could be contaminated by overlying pits and shallow OGD wells. Risk of this type of contamination scenario also depends on geological characteristics. Among all OGD well types, we were only able to obtain the depth for HF wells from their permit records; the "All Wells" dataset did not provide data on depth. The domestic well areas provide a general area where domestic wells are operating without any information about the wells' specific location and depth. Similarly, while we can identify the CWS with groundwater as the primary source, we do not know the location of wells that supply water to drinking water systems. California's Groundwater Information System (GAMA) provides data on well depth from domestic wells for various purposes (observation, industrial, irrigation, residential, stockwatering, other and unknown) as well as TDS levels (SWRCB 2019). GAMA wells did not overlap with all DWA and CWS in our analysis and averaging depth and TDS from available wells would not be an accurate representation of individual domestic wells or CWS wells. For example, industrial and residential wells are not equivalent as an industrial well may be tapping into an area with higher TDS and greater depths than a residential well within the same block group or CWS boundary. Only 12% of GAMA wells are residential wells. Due to these limitations, we were unable to further characterize whether contamination from OGD infrastructure could significantly impact local drinking water users within the area. USGS's preliminary groundwater salinity analysis with historical data indicates that TDS is the highest and found at the shallowest depths within the south-western region of the SJV (Metzger and Landon 2018), where several CWS intersect with OGD and an exempt aquifer.
Another limitation of our analysis was that we were unable to tap into violations data to identify, for example, those pits that received cease and desist orders or oil and gas wells that had a history of spills and leaks. Information on the violations status and type would have to combined with water sample data (if it exists) to track the source and transport of OGD contaminants within water systems to better characterize the extent and specificity of potential threats on drinking water resources from OGD. Water transport models would also have to be developed to trace contamination from spills and leaks. This is beyond the scope of our analysis.

In summary, we identified that 12% of block groups with domestic well areas and community water systems are at risk of water contamination from OGD in SJV. All four well types (OGP, HF, EOR, WD) and two unlined pit types (active and inactive) were found within 3 km of DWA while OGP, EOR, WD and inactive unlined pits were found within 1 km of CWS. No SES or area-level factors significantly predicted the overall number of OGD wells and pits per km^2 except CWS system size. However, the effect estimates suggest potentially informative relationships. For example, block groups with a lower proportion of Hispanics predicted higher counts of OGD while CWS with a higher proportion of Hispanics predicted higher counts of OGD. Our results demonstrate that different types of water systems in an area can be vulnerable in different ways as SES factors and the type and volume of OGD infrastructure vary throughout the region. Additionally, small CWS are much more vulnerable to contamination threats than larger systems. Future studies could expand upon our work by considering other SES predictors or integrating violations data, and conducting similar analyses on water systems in other states with OGD. Our work highlights the need for increased water sampling and monitoring efforts, particularly within small water systems. Without these critical water quality sampling data, it is challenging to establish a baseline to determine whether OGD is impacting drinking water sources, even without major spills or leaks. Regulatory agencies might also consider applying the same exemption criteria across all OGD wells (production and injection) and pits to expand the protection of water sources that could be used for drinking across the state.

3.6 Tables

Table 3.1. Population and area level characteristics of block groups with domestic well areas that were included and excluded in the analysis.

Measures	**BG without OGD** *n=431*	**BG with OGD** *n=58*	**Excluded BG** *n=3*
Total num of domestic wells	38,832	5,421	118
Mean num of domestic wells	90	93	39
Total population served	233,034	39,717	2,541
Mean population served	541	685	847
Mean domestic well area (km^2)	42	109	127
Mean % DWA at risk	0	26	34
Mean DWA population at risk	0	137	314
Mean population density (people/km^2)	63	24	770
Mean % 2x below poverty	45	50	42
Mean % Hispanics	49	59	21
Num of BG's with shale play (%)	9 (2)	10 (17)	1 (33)
Mean OGD count	0	41	2,948

Note: BG, block group; OGD, oil and gas development infrastructure; DWA, domestic well area.

Table 3.2. Adjusted estimates and exponentiated estimates by GLM and GAM negative binomial models for block groups with domestic well areas.

	GLM negative binomial		*GAM negative binomial*	
Predictors	**Estimate (95% CI)**	**Exp(Est) (95% CI)**	**Estimate (95% CI)**	**Exp(Est) (95% CI)**
% Poverty[a]	0.26 (-0.29, 0.79)	1.3 (0.75, 2.21)	0.014 (-0.33, 0.36)	1.01 (0.72,1.43)
% Hispanic[a]	0.017 (-0.44, 0.50)	1.02 (0.64, 1.65)	-0.098 (-0.42, 0.23)	0.91 (0.66, 1.25)
Shale play (presence)	-0.59 (-2.88, 4.07)	0.55 (0.056, 58.6)	1.86 (-0.51, 4.22)	6.41 (0.60, 68.2)

Note: GLM, generalized linear model; GAM, generalized additive model; Exp, exponentiated; Est, estimate. Model adjusted for population density.
[a]Per 10% increase.

Table 3.3. Population and area level characteristics of community water systems that were included and excluded in the analysis.

Measures	**CWS without OGD** *n=376*	**CWS with OGD** *n=36*	**Excluded CWS** *n=5*
Total num of service connections	243,974	45212	18,278
Mean num of service connections	649	1,256	3,656
Total population served	754,101	163,999	47,334
Mean population served	2,006	4,556	9,467
Mean population density (people/km^2)	817	645	146
Mean % 2x below poverty	42	37	50
Mean % Hispanics	39	47	34
Num of CWS's with shale play (%)	10 (3)	4 (11)	2 (40)
Small system count (%)	281 (75)	28 (77)	2 (40)
Intermediate system count (%)	80 (21)	2 (6)	1 (20)
Medium system count (%)	11 (3)	4 (11)	2 (40)
Large system count (%)	4 (1)	2 (6)	0 (0)
Mean OGD count	0	34	6,533

Note: OGD, oil and gas development infrastructure; CWS, community water system.

Table 3.4. Adjusted estimates and exponentiated estimates from the generalized linear negative binomial model for community water systems.

	GLM negative binomial	
Predictors	**Estimate (95% CI)**	**Exp(Est) (95% CI)**
% Poverty[a]	-0.44 (-1.00, 0.11)	0.64 (0.37, 1.12)
% Hispanic[a]	0.32 (-0.16, 0.79)	1.37 (0.86, 2.20)
Shale play (presence)	0.19 (-3.70, 4.07)	1.21 (0.025, 58.6)
CWS size - small	5.41 (3.10, 7.72)	223 (22.1, 2247)

Note: Model adjusted for log(population density).
[a]Per 10% increase.

3.7 Figures

Figure 3.1. Map of active oil and gas wells and unlined active and inactive percolation pits relative to domestic well areas within the study region of San Joaquin Valley, California.

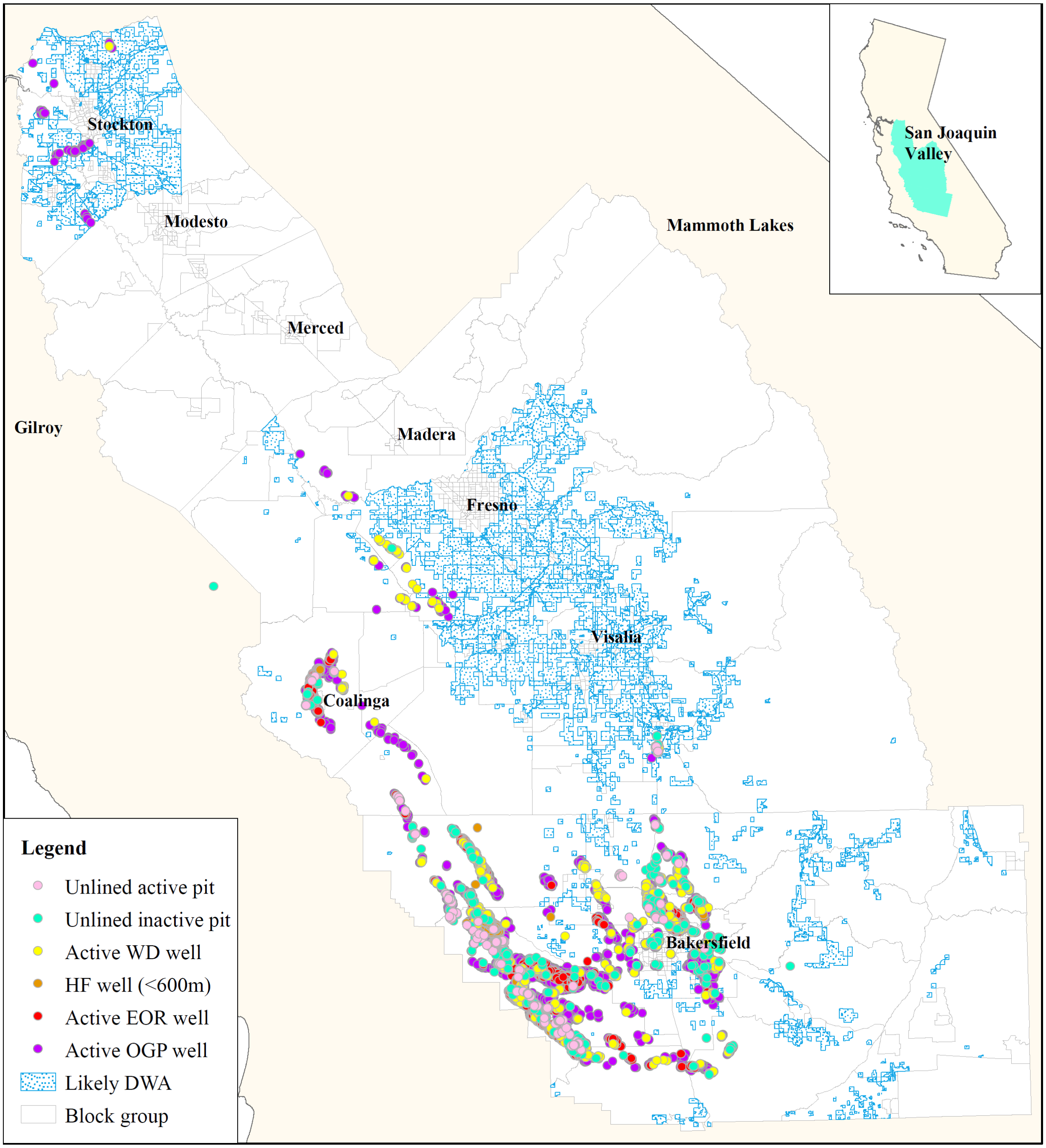
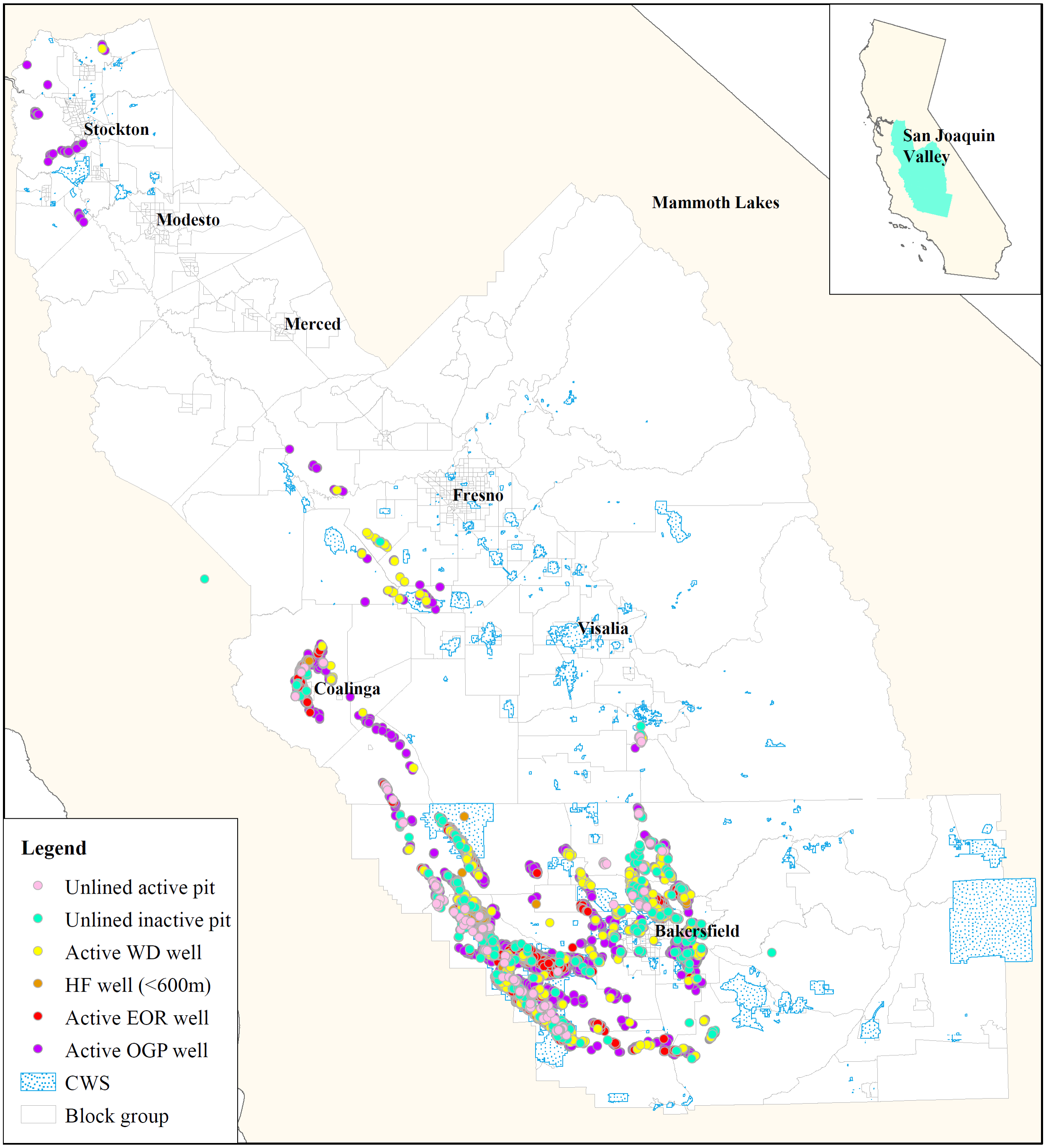


Figure 3.2. Map of active oil and gas wells and unlined active and inactive percolation pits relative to community water systems within the study region of San Joaquin Valley, California.



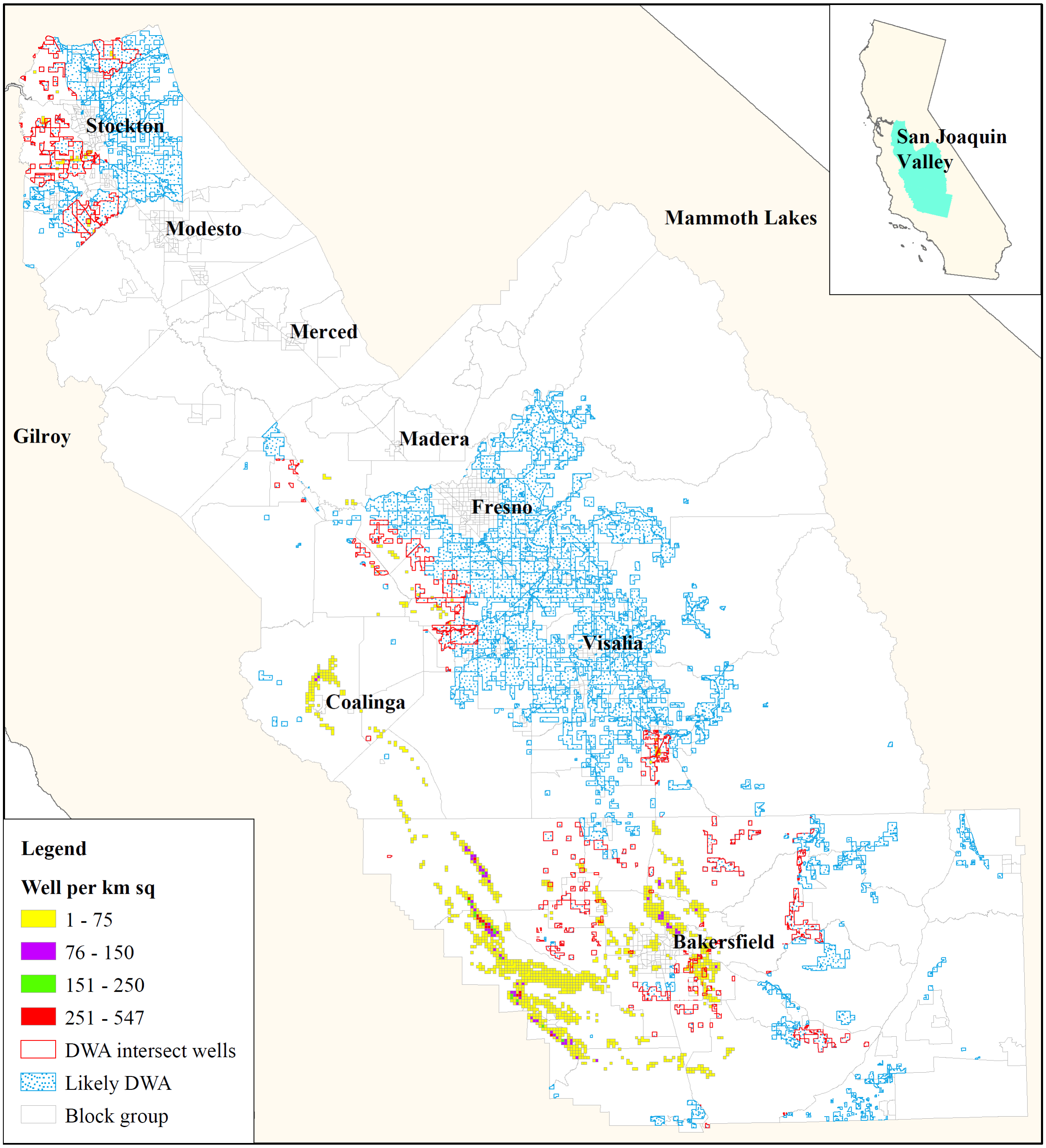
3.8 Supplemental Information Chapter 3

Supplemental Table 3.1. Diagnostic parameters for regression models.

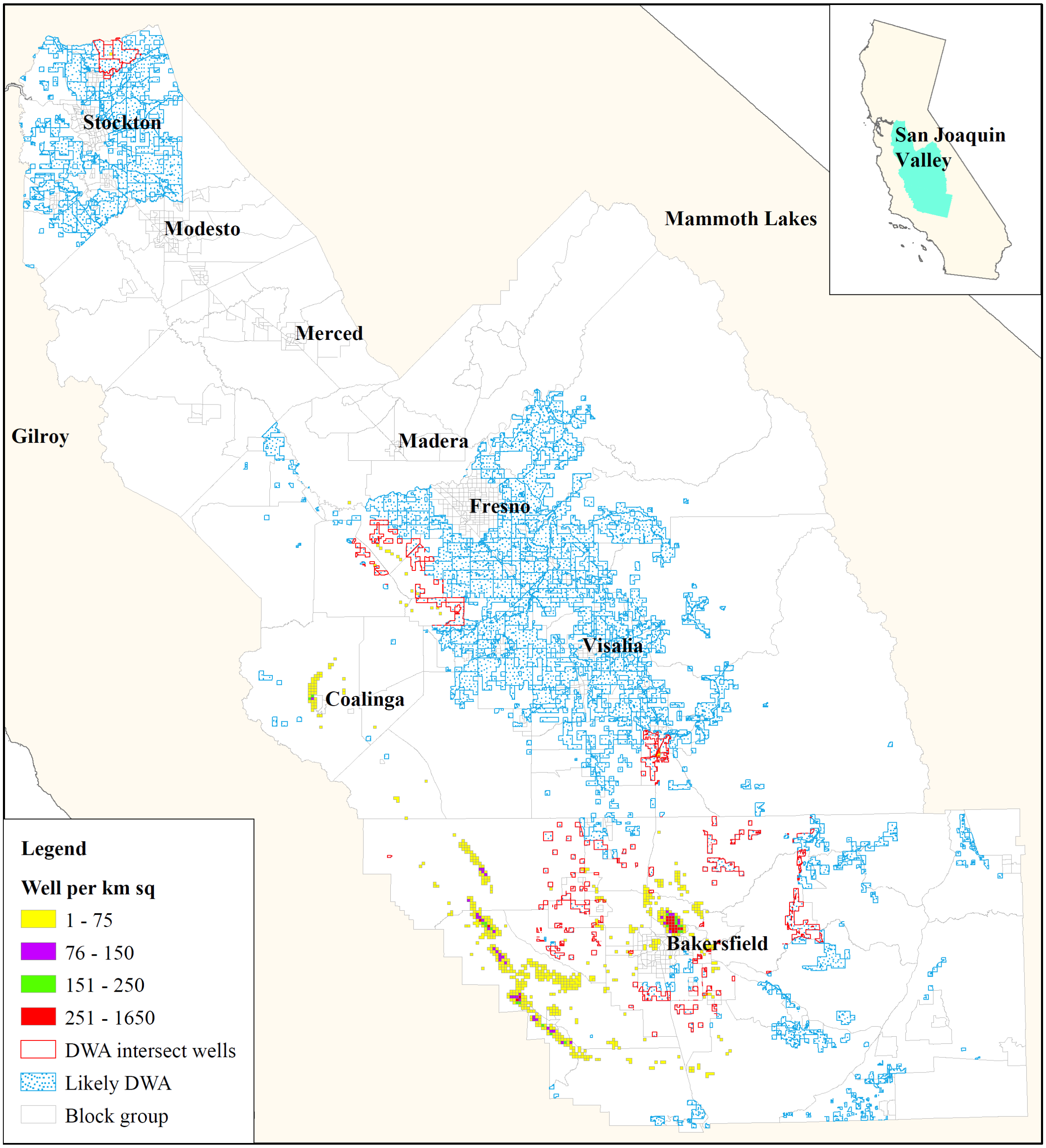
Model	Water		Dispersion	Moran's I
GLM Poisson	DWA	14,314	280.42	<2.2 x10^{-16}
GLM negative binomial	DWA	860	0.89	<2.2 x10^{-16}
GAM negative binomial	DWA	579	0.089	1
GLM Poisson	CWS	7763	119.84	0.4618
GLM negative binomial	CWS	573	0.95	0.3117

Note: DWA, domestic well area; CWS, community water system; GLM, generalized linear model; GAM, generalized additive model.

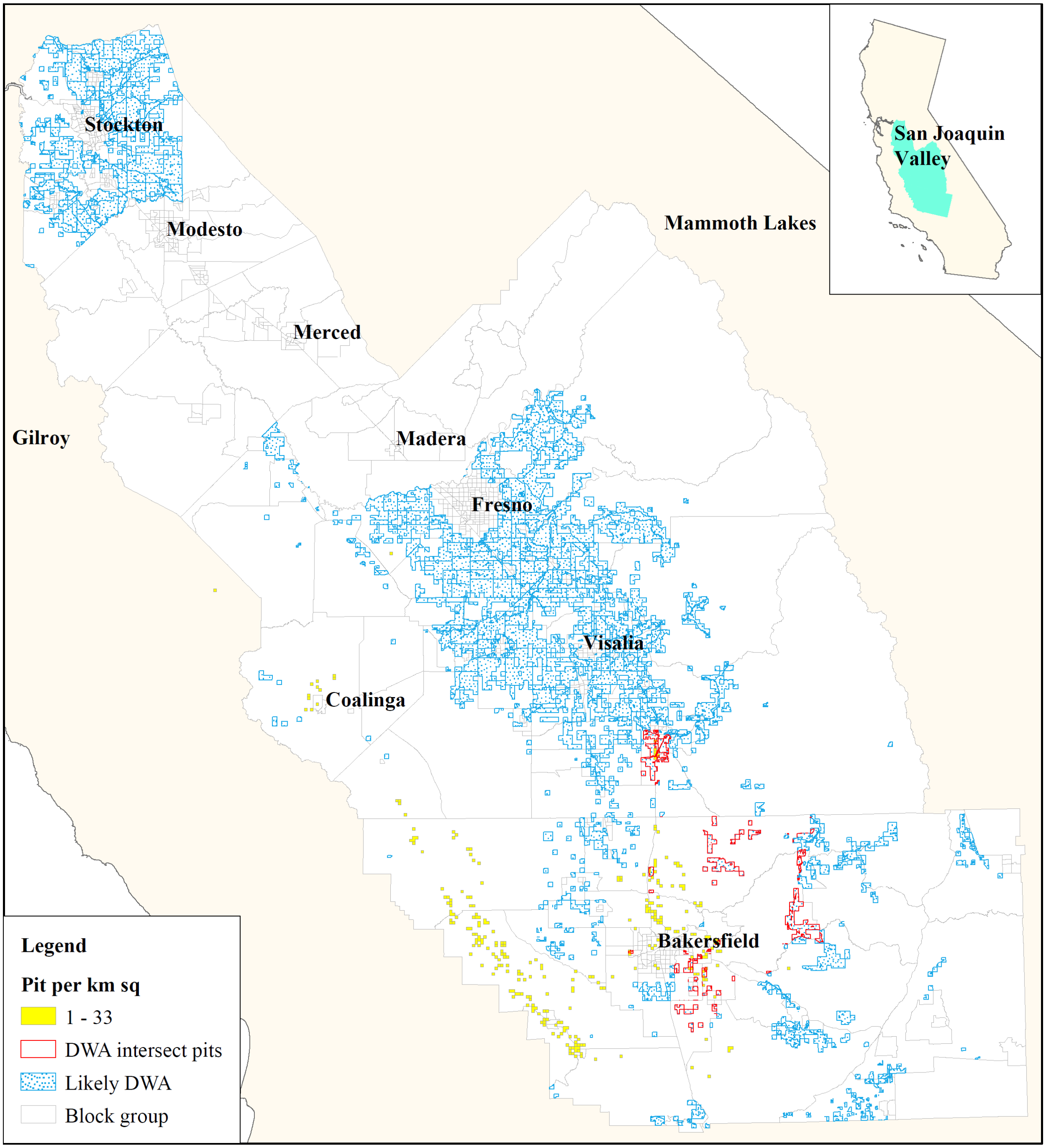
Supplemental Figure 3.1. Map of domestic well areas (DWA) and density (wells per km2) of oil and gas production wells and shallow (<600 m) hydraulically fractured wells within the study region of San Joaquin Valley, California.



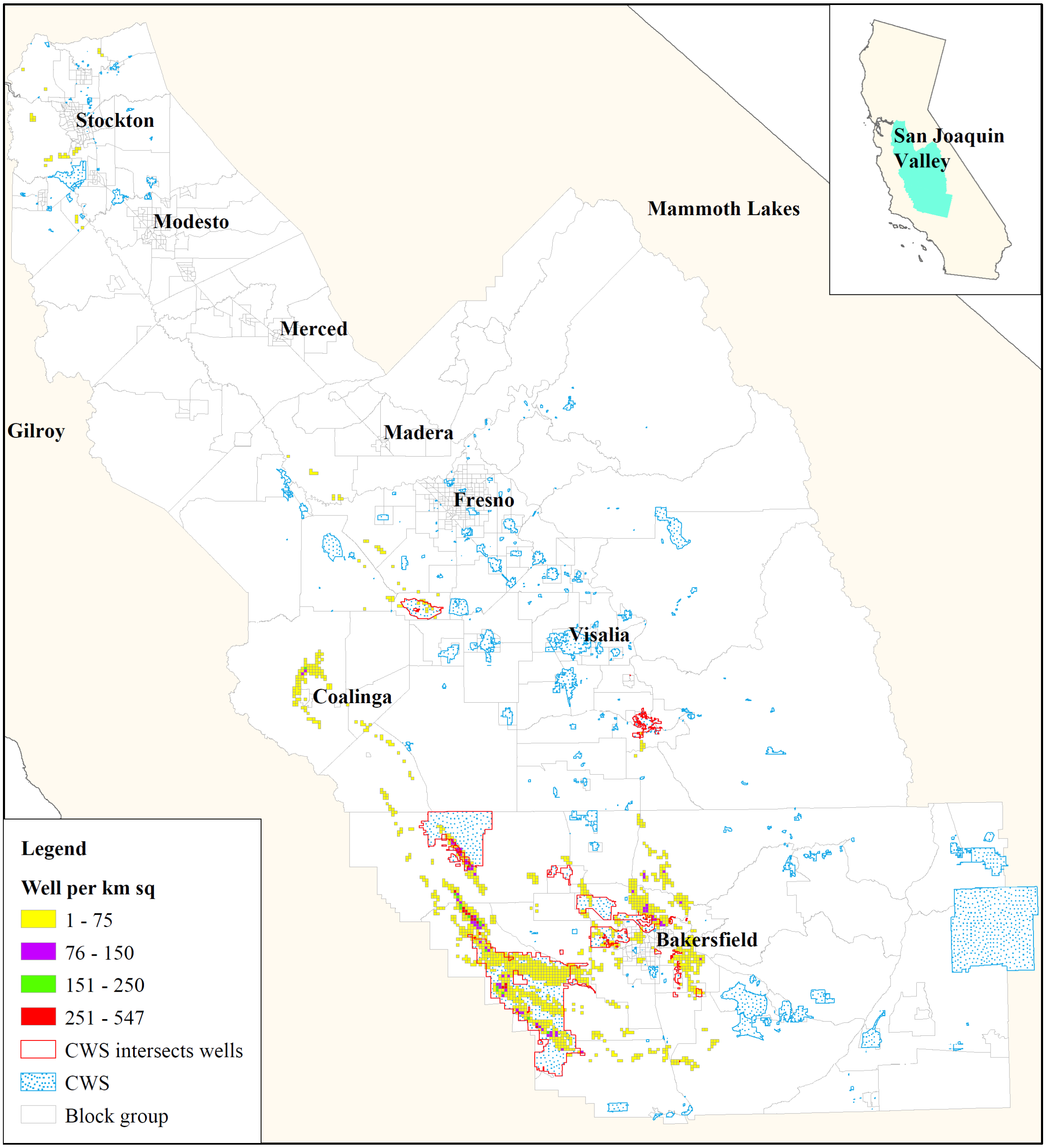
Supplemental Figure 3.2. Map of domestic well areas (DWA) and density (wells per km2) of injection wells (enhanced oil recovery and waste disposal wells) within the study region of San Joaquin Valley, California.



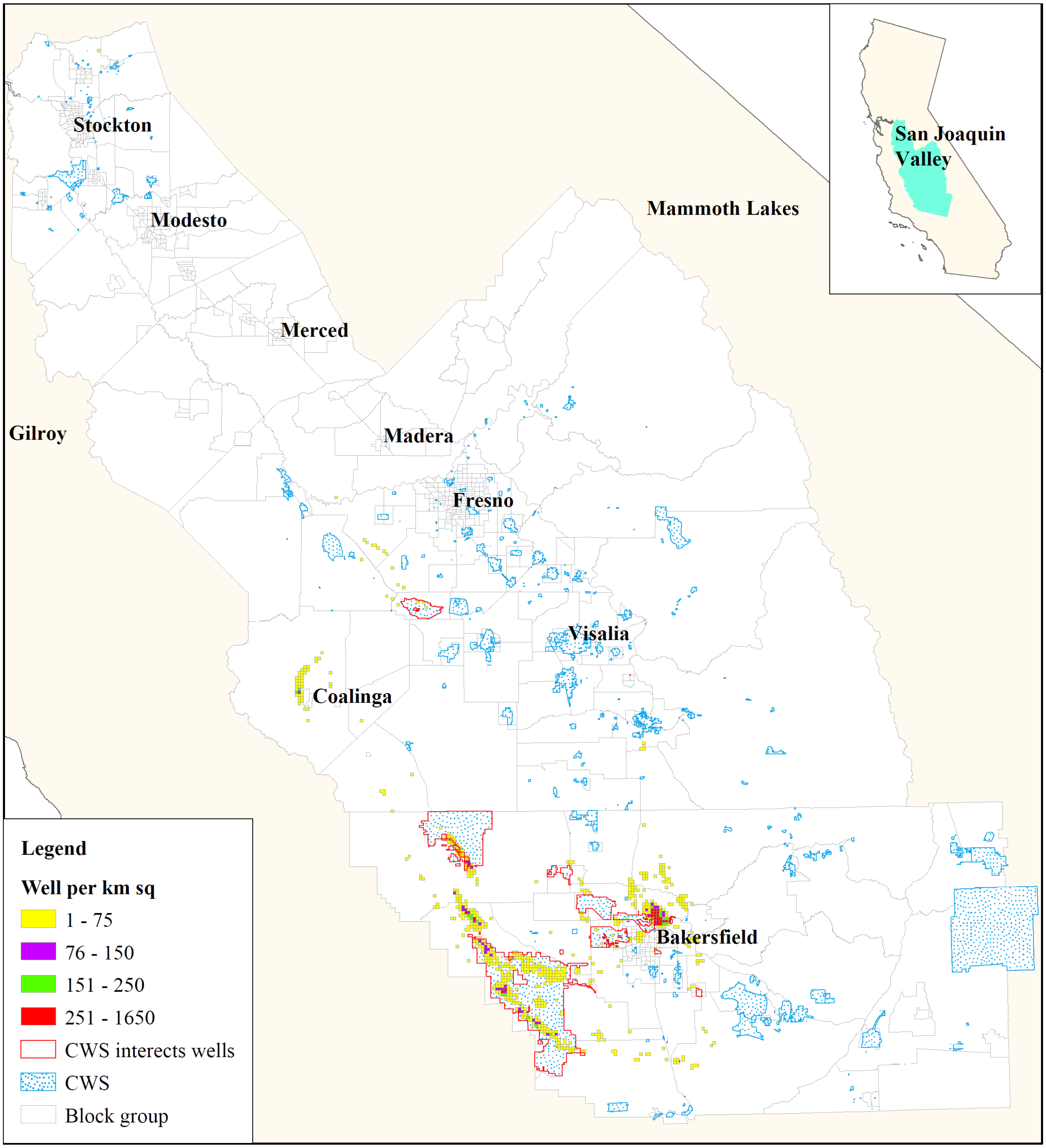
Supplemental Figure 3.3. Map of domestic well areas (DWA) and density (pits per km^2) of active, inactive and closing percolation pits within the study region of San Joaquin Valley, California.



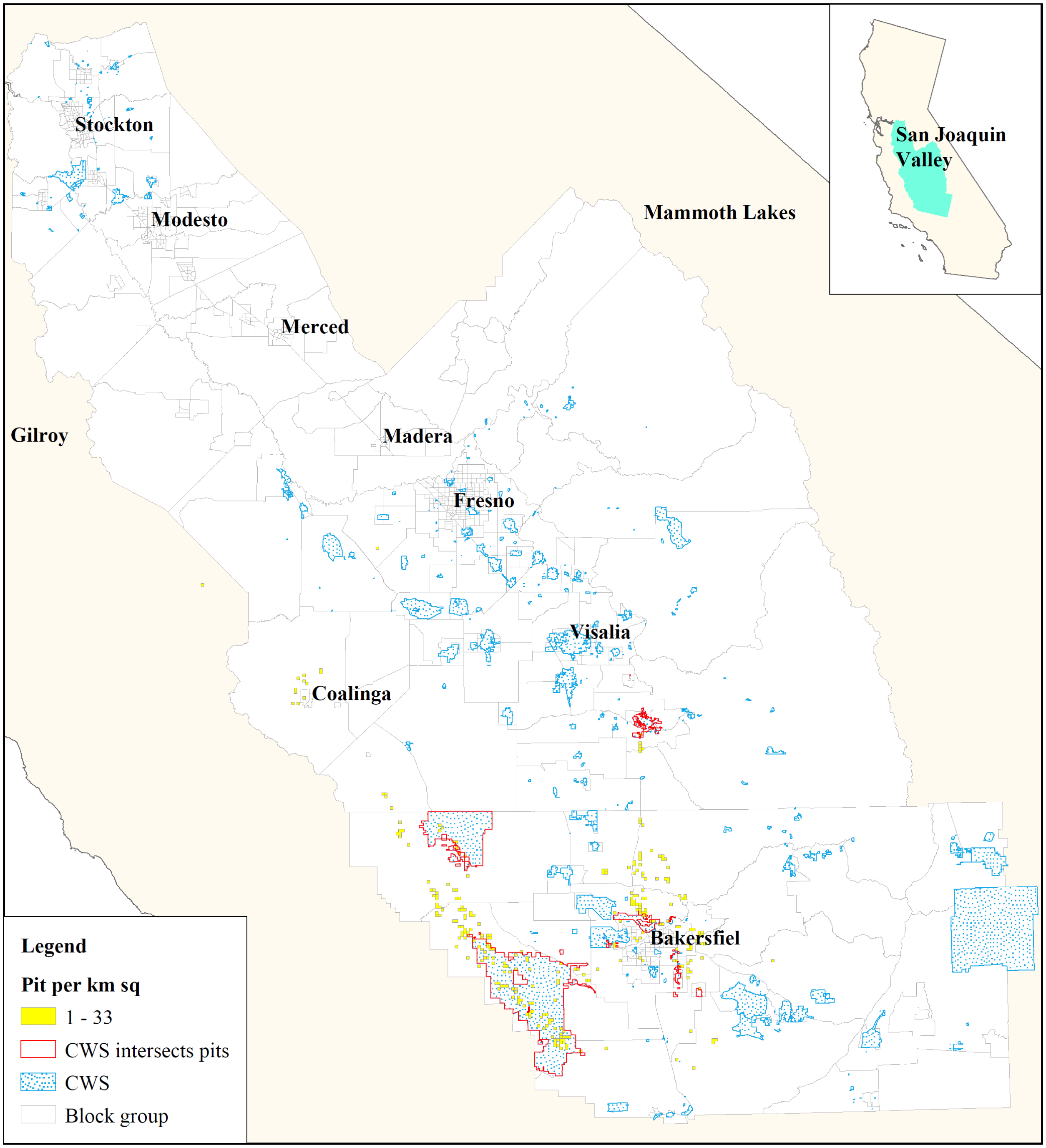
Supplemental Figure 3.4. Map of community water systems (CWS) and density (wells per km^2) of oil and gas production wells and shallow (<600 m) hydraulically fractured wells within the study region of San Joaquin Valley, California.



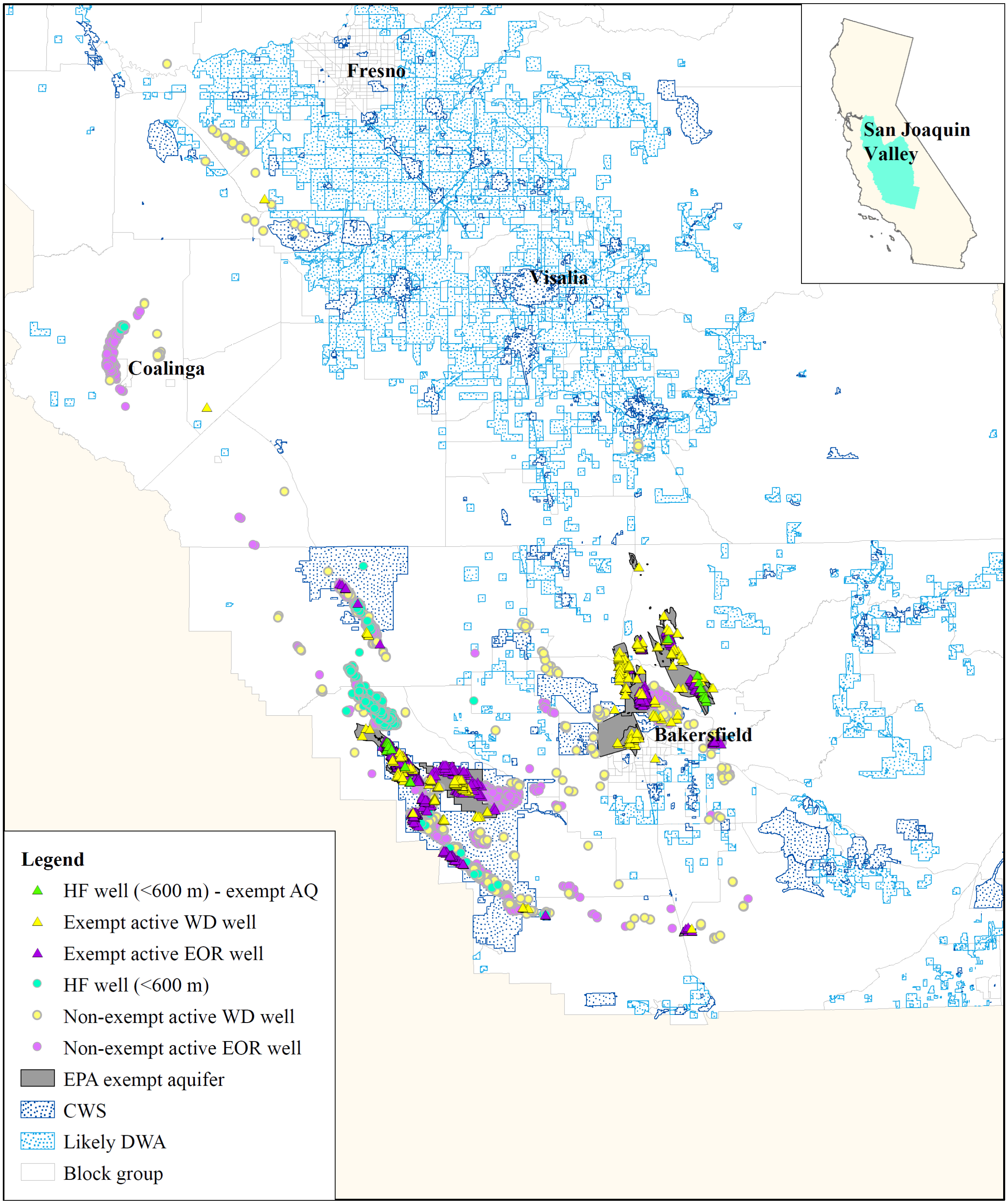
Supplemental Figure 3.5. Map of community water systems (CWS) and density (wells per km^2) of injection wells (enhanced oil recovery and waste disposal) within the study region of San Joaquin Valley, California.



Supplemental Figure 3.6. Map of community water systems (CWS) and density (pits per km^2) of active, inactive and closing percolation pits within the study region of San Joaquin Valley, California.



Supplemental Figure 3.7. Map of aquifers (AQ) with exemption status from the US EPA and exemption status of Class II injection wells (based on designation by CalGEM) and overlap with exempt aquifers, shallow hydraulically fractured wells (based on overlap with exempt aquifers), and unlined pits (based on overlap with exempt aquifers).



Supplemental Table 3.2. Mean, median and range of counts for each oil and gas well and percolation pit type by water system and grouped by whether block group or CWS were included or excluded from analyses.

	Block Group (DWA)				*CWS*			
Measures	**n**	**Mean**	**Median**	**Range**	**n**	**Mean**	**Median**	**Range**
Analyzed observations								
Count of all OGD types	58	41	12.5	1 - 326	36	34	21.5	1 - 275
Count of OG production wells	58	28	6.5	1 - 316	36	21	15.5	1 - 105
Count of HF well (<600 m)	2	2.5	2.5	1 - 4	0	0	0	0
Count of EOR wells	11	40	7	1 - 170	4	87	41.5	2 - 263
Count of WD wells	30	4	3	1 - 21	22	3	2	1 - 8
Count of active percolation pits	6	10	10	1 - 17	0	0	0	0
Count of inactive percolation pits	19	7	4	1 - 29	16	3.25	4	1 - 6
Excluded observations (outliers)								
Count of all OGD types	3	2,948	3,789	620-4,435	5	6,533	3,598	1,897 - 19,185
Count of OG production wells	3	632	559	453-884	5	3,169	1,128	261 - 11,850
Count of HF well (<600 m)	2	2.5	2.5	1-4	3	53	69	1 - 90
Count of EOR wells	3	2,277	3,208	120-3,504	5	3,129	1,773	1,123 - 6,416
Count of WD wells	3	27	28	13-40	5	90	30	11 - 353
Count of active percolation pits	3	2.7	2	1-5	2	61	0	14 - 290
Count of inactive percolation pits	3	7	6	3-13	5	53	13	6-207
Count of all OGD types	3	2,948	3,789	620-4,435	5	6,533	3,598	1,897 - 19,185

Note: DWA, domestic well area; CWS, community water system; OGD, oil and gas development, OG, oil and gas; HF, hydraulic fracturing; EOR, enhanced oil recover; WD, waste disposal.

Chapter 4: Air pollution, high methane emitters, and oil and gas wells in Northern California: the relationship with migraine headache prevalence and exacerbation

Publication: Elser H, Morello-Frosch R, Jacobson A, Pressman A, Kioumourtzoglou MA, Reimer R, Casey JA (2021) Air pollution, methane super-emitters, and oil and gas wells in Northern California: the relationship with migraine headache prevalence and exacerbation. *Environmental Health* 20:45. https://doi.org/10.1186/s12940-021-00727-w

4.1 Abstract

Background: Migraine–an episodic disorder characterized by severe headache that can lead to disability–affects over 1 billion people worldwide. Prior studies have found that short-term exposure to fine particulate matter ($PM_{2.5}$), nitrogen dioxide (NO_2), and ozone increases risk of migraine-related emergency department (ED) visits. Our objective was to characterize the association between long-term exposure to sources of harmful emissions and common air pollutants with both migraine headache and, among patients with migraine, headache severity.
Methods: From the Sutter Health electronic health record database, we identified 89,575 prevalent migraine cases between 2014–2018 using a migraine probability algorithm (MPA) score and 270,564 frequency-matched controls. Sutter Health delivers care to 3.5 million patients annually in Northern California. Exposures included 2015 annual average block group-level $PM_{2.5}$ and NO_2 concentrations, inverse-distance weighted (IDW) methane emissions from 60 high-emitters located within 10km of participant residence between 2016–2018, and IDW active oil and gas wells in 2015 within 10km of each participant. We used logistic and negative binomial mixed models to evaluate the association between environmental exposures and (1) migraine case status; and (2) migraine severity (i.e., MPA score >100, triptan prescriptions, neurology visits, urgent care migraine visits, and ED migraine visits per person-year). Models controlled for age, sex, race/ethnicity, Medicaid use, primary care visits, and block group-level population density and poverty.
Results: In adjusted analyses, for each 5ppb increase in NO_2, we observed 2% increased odds of migraine case status (95% CI: 1.00, 1.05) and for each 100,000 kg/hour increase in IDW methane emissions, the odds of case status also increased (OR = 1.04, 95% CI: 1.00, 1.08). We found no association between $PM_{2.5}$ or oil and gas wells and migraine case status. $PM_{2.5}$ was linearly associated with neurology visits, migraine-specific urgent care visits, and MPA score >100, but not triptans or ED visits. NO_2 was associated with migraine-specific urgent care and ED visits, but not other severity measures. We observed limited or null associations between continuous measures of methane emissions and proximity to oil and gas wells and migraine severity.
Conclusions: Our findings illustrate the potential role of long-term exposure to multiple ambient air pollutants for prevalent migraine and migraine severity.

4.2 Introduction

Migraine is an episodic disorder characterized by severe headache often associated with nausea or sensitivity to light and sound. In 2016, the estimated global prevalence of migraine was 14.4% with over 1.04 billion individuals affected worldwide (Stovner et al. 2018). In the United States (U.S.), migraine is most common among individuals aged 30 to 39 and follows a social gradient wherein migraine is less common among wealthier individuals (Lipton et al. 2001a, 2007). Migraine can lead to disability; in the U.S., estimated annual costs associated with migraine range from $13 to 16.6 billion annually due to lost productivity, work and school absences, and short-term disability (Berg and Ramadan 2006; Gilligan et al. 2018; Lofland 2007; Porter et al. 2019).

Given the episodic nature of migraine headache, considerable attention has been paid to the study and identification of common triggers. Among the most frequently self-reported triggers of migraine are sleep disturbances and fatigue; stress or relief of stress; menstruation and pregnancy; smoking; and food and alcohol (Chabriat et al. 1999; Henry et al. 2002; Peatfield et al. 1984; Prince et al. 2004; Spierings et al. 2001). Factors such as noise, season, and weather variations have also been implicated as migraine triggers (Charles 2013; Eross et al. 2007; Prince et al. 2004; Wöber et al. 2006). Examples of common sources of environmental noise that may precipitate a migraine attack include traffic-related noise from roads, railways, aircrafts, and parking cars (Friedman and De Ver Dye 2009). Individuals with migraine frequently attribute their headaches to weather variations, including changes in temperature and barometric pressure (Spierings et al. 2001; Turner et al. 1995; von Mackensen et al. 2005; Wöber et al. 2006; Yang et al. 2011).

Research to date also implicates short-term exposure to a variety of air pollutants as triggers for migraine headache. Fine particulate matter ($PM_{2.5}$) is among the most frequently studied pollutants; increased levels of $PM_{2.5}$ have been associated with more frequent migraine-specific emergency department (ED) visits in Canada, Taipei, and South Korea (Chen et al. 2015; Chiu et al. 2015; Lee et al. 2018; Szyszkowicz et al. 2009a, 2009c), although, a case-crossover study of 7,054 patients in Boston reported no significant association with ED visits (Mukamal et al. 2009). In a time-series study of 1,059 ED visits recorded at a Vancouver hospital, levels of sulfur dioxide (SO_2) were associated with ED visits for migraine (Szyszkowicz et al. 2009b). Levels of ozone, carbon monoxide, nitrogen dioxide (NO_2), and coarse particles (PM_{10}) have also been linked with migraine-specific ED visits in case-crossover studies based on daily clinic data from 1,000,000 patients from the National Health Insurance Program in Taiwan (Chen et al. 2015; Chiu et al. 2015). A cross-sectional survey of 7,785 primary care patients of the Geisinger Clinic in 2014 found that individuals exposed to the highest levels of unconventional natural gas development were more likely to have migraine headache (Tustin et al. 2017b). Unconventional natural gas development can produce $PM_{2.5}$, volatile organic compounds (VOCs), noise and light pollution, and stressful community changes that could trigger migraine (Adgate et al. 2014c).

To date, few studies have considered the implications of long-term exposure to common environmental pollutants – which may capture potential residential disparities in the burden of headache based on local average air quality – and no analyses have been conducted in the Western U.S. or on specific air pollution sources. Recently, the California Air Resources Board (CARB) conducted an air survey of high methane emitters, point sources of methane emissions, including dairies, landfills, refineries, and oil and gas infrastructure (Duren et al. 2019a). These facilities emit a variety of co-pollutants such as SO_2, hydrogen sulfide, $PM_{2.5}$, and VOCs (California Air Resources Board 2019, 2020; Zavala-Araiza et al. 2017), and the new CARB data provide an opportunity to assess their implications for migraine. California is also a top-10 U.S. producer of crude oil, with over 200,000 oil and gas wells drilled in the state (Energy Information Agency 2020).

The present study leverages data from the Sutter Health electronic health record (EHR) database in Northern California and builds on prior research linking air pollutants and migraine headache. Our analyses include an expanded set of exposure measures, including long-term ambient $PM_{2.5}$ and NO_2 concentrations, methane emissions, and active oil and gas wells as measured at the beginning of the study period. We selected these exposures, particularly high methane emitters and active oil and gas wells, because if found to be linked to migraine, policies could reduce emissions at the source. Whereas past research has largely relied on migraine-specific ED visits as a crude proxy for severe headache, we incorporate additional measures of headache severity. We conducted a case-control study to ascertain whether migraine case status was associated with long-term exposure to any of the four environmental exposure measures as compared with controls. Next, we conducted a case-case analysis to ascertain whether environmental exposures were associated with more severe headache among individuals with established diagnosis of migraine. We hypothesized *a priori* that environmental exposures would be associated with both migraine case status and with disease severity.

4.3 Methods

We conducted a case-control study and case-case analysis to examine the relationship between migraine severity and exposures of interest. This approach was selected based on computational feasibility, and because disease-based sampling is efficient when multiple exposures are considered and when the outcome of interest is relatively rare (Jewell 2003). Cases and controls were identified through the Sutter Health EHR database. Sutter Health is a large, mixed-payer, integrated healthcare system in Northern California that delivers comprehensive medical services through its network of 24 acute-care hospitals and more than 100 ambulatory clinics. Approximately 3.5 million patients receive care through Sutter each year at hospitals and clinics located in 22 counties; our study subjects resided in 27 urban and rural counties. Sutter's Epic EHR (Epic Systems Corporation, Verona, Wisconsin) is fully integrated across all hospital and ambulatory sites. Data for cases and controls were retrospectively extracted from the Sutter EHR for the study period between January 1, 2014 and December 31, 2018.

Patient demographic data from the EHR included sex (male, female), race/ethnicity (non-Hispanic Asian, Black, white, other, or Hispanic); and marital status (divorced, separated, widowed; married or partnered; single; other or unknown). We used date of birth to compute age in years at the start of follow-up. Health characteristics extracted from the EHR included whether the individual was a Medicaid beneficiary (yes, no); body mass index (BMI) category in kg/m^2 [less than 18.5 (underweight); 18.5 – 24.9 (normal); 25 – 29.9 (overweight); 30 – 34.9 (obese class 1); 35 – 39.9 (obese class 2); 40 or more (obese class 3)]; number of and reason for primary care, specialty care, urgent care, and emergency department visits. We assigned residential address for the study period (2014–2018) based on address of record in October 2019. Using assigned residential address, we linked block group-level percent living below the federal poverty threshold and population density (individuals per km^2) using data from the 2014–2018 American Community Survey.

Migraine Case Ascertainment and Control Selection

Both cases and controls were selected from the study base of eligible patients over the age of 18 with at least one primary care encounter during the five-year study period (2014–2018) that resided in one of 27 counties in Northern California. We ascertained case status using the Migraine Probability Algorithm (MPA), a validated approach for identification of individuals diagnosed with migraine from EHR data (Pressman et al. 2016). Briefly, a numeric score that ranges from zero to 101 is calculated based on the following criteria: encounters (hospital inpatient, emergency room and outpatient) with a primary or secondary diagnostic code for migraine from the *International Classification of Diseases*, Ninth Revision (*ICD*-9 346.xx) or Tenth Revision (*ICD*-10 G43.xxx); an *ICD*-9 or *ICD-10* code for migraine in the patient's Significant Health Problem List (SHP); and filled prescriptions for migraine-specific abortive medications (i.e., triptans, ergotamines). An MPA score greater than 10 is consistent with diagnosis of migraine. We selected three controls for every case from the Sutter EHR database. Controls were frequency matched to cases based on age category (18–29; 30–44; 45–54; 55–64; 65 or older), sex, year of entry into Sutter primary care, and primary-care follow-up time (0–6 months, 7–24 months, > 24 months).

Migraine Severity

Among cases (i.e., individuals with MPA > 10), we defined the following count variables to capture migraine severity (1) all-cause neurology visits per year; (2) migraine-specific urgent care visits per year; (3) triptans prescribed per year. We additionally defined two dichotomous measures to capture migraine severity: (4) 0 versus ≥ 1 migraine-specific emergency department (ED) visit during the study period; and (5) MPA score > 100 (more severe) versus MPA score 11-100 (less severe).

Air pollution, methane emission, and oil and gas wells

We considered four separate exposure measures in our analyses. These included $PM_{2.5}$, NO_2, high methane emitters, and active oil and gas wells. Exposure to air pollutants and to oil and gas wells was estimated based on average values at the beginning of the study period (in 2015). Methane emissions measures were based on data collected between 2016–2018.

$PM_{2.5}$ and NO_2

We used patient addresses to link annual average concentration of $PM_{2.5}$ and NO_2 estimates at the block group-level derived from annual-average integrated empirical geographic regression models (Kim et al. 2020b). The approach relied on universal kriging and took regulatory monitoring data, satellite imagery, and measures of land use and traffic as inputs. $PM_{2.5}$ and NO_2 achieved standardized RMSEs of 0.86μg/m^3 and 0.87ppb, respectively. These variables were re-scaled such that coefficients in linear models correspond to each 5μg/m^3 increase in $PM_{2.5}$ and each 5ppb increase in NO_2, respectively.

Methane Emissions

Data on methane emissions were provided by CARB as described in Duren et al. 2019 (Duren et al. 2019a). In brief, CARB led the first California Methane Survey to provide systematic information on methane point sources across the state via Next Generation Airborne Visible/Infrared Imaging Spectrometer (AVIRIS-NG) flights conducted between 2016–2018. The AVIRIS-NG flights identified 564 distinct sources of methane plumes and captured average hourly emission rates in kilograms per hour (kg/hour). Examples of sources of methane plumes identified by AVIRIS-NG flights included oil and gas wells, dairies, and landfills. To estimate exposure to methane emissions for the present study, we calculated the sum kg/hour of emitted methane from all sources within 10km of each participant *j*'s residence and weighted the emissions by the inverse-distance squared between each high methane emitter, *i*, and patient *j*'s residence:

$$\sum_{i=1}^{n} \frac{E_i}{d_{ij}^2},$$

where *E* is the emission rate at high methane emitter *i* in kg/hour and *d* is the distance in kilometers between high methane emitter *i* and participant *j*. We created two exposure metrics based on methane emission rates. The first was the sum of methane emissions (in kg/hour) within 10km, re-scaled so that model coefficients corresponded to a 100,000 kg/hour increase in methane emissions. The second was an indicator variable for presence of any high methane emitter within 10km.

Oil and Gas Wells

Finally, we obtained records for active oil and gas wells as of December 2015 from the California Division of Oil, Gas and Geothermal Resources website (CA DOGGR). To estimate exposure to active wells, we used inverse-distance weighting (IDW) of active wells within 10km of each participant, *j*:

$$\sum_{i=1}^{n} \frac{1}{d_{ij}^2},$$

where *i* is an active well located within 10km of the participant and *d* is the distance in kilometers between well *i* and participant *j*. We created two exposure metrics based on exposure to active wells. The first was a continuous IDW sum of all active wells within 10km, re-scaled so that coefficients in linear models correspond to a 1,000-unit increase in the IDW sum. The second was an indicator variable for presence of any active oil or gas well within 10km.

Statistical Analyses

We first conducted a case-control analysis in which we examined the association between migraine status and each of the four exposures. Next, we conducted a case-case analysis to examine whether migraine severity was associated with each of the exposures.

Case-Control Analysis

For the case-control analyses, we used generalized linear mixed models with a logit link with county-specific random intercepts to account for potential within-county clustering. All models controlled for our matching variables: categorical age and sex, as recommended (Mansournia et al. 2018), and race/ethnicity, Medicaid use, number of primary care visits per year, and block group-level population density and poverty. We specified four separate statistical models to examine the association between migraine status and each of the environmental exposures of interest (i.e., $PM_{2.5}$, NO_2, high methane emitters, and active oil and gas wells). We used generalized additive mixed models with penalized smoothing splines to capture potential non-linearities in the exposure-response relationships. As a secondary analysis, we used the binary exposure specification for both high methane emitters and active wells (i.e., any high methane emitter within 10km vs. none and any well within 10km vs. none).

Case-Case Analysis
In the case-case analyses, we utilized negative binomial mixed models (for count of neurology visits, migraine-related urgent care visits, and prescriptions for triptans) and logistic mixed models (for ≥1 migraine-related ED visit per person-year vs. less and MPA score >100 vs. 10-100) with random intercepts for county to examine the association between migraine severity and the exposure of interest. We controlled for the same set of potential confounding variables as described for the case-control analysis, assessed deviations from linearity using penalized smoothing splines, and as a secondary analysis considered binary specifications of high-emitters and active wells.

Sensitivity Analyses
We conducted the following sensitivity analyses. First, we separated exposure to high-emitters into two categories: (1) dairy/cattle manure and landfills and (2) all other industrial types, which included power plants, refineries, wastewater treatment facilities, oil and gas distribution (e.g., oil/gas compressors, gas distribution lines), and oil and gas production (e.g., oil/gas waste lagoons, oil/gas plugged wells). We did so under the assumption that methane co-pollutant emissions would differ by these two categories. Second, we repeated our main case-control and case-case analyses with additional adjustment for BMI category and marital status. Finally, in our case-case analysis of migraine-specific ED visits, we additionally adjusted for distance from the patient's address of record to the nearest ED.

For all models, we evaluated residual spatial autocorrelation using Moran's I (Bivand 2008), which indicated no residual spatial autocorrelation in any of the analyses. Analyses were conducted using R 3.6.0 (R Foundation for Statistical Computing, Vienna, Austria). The Columbia University (Protocol #: AAAT0085), University of California, Berkeley (Protocol #: 2013-10-5693), and Sutter Health (IRBNet #:1452543-1) Institutional Review Boards approved this study.

4.4 Results

The study based included 1,433,236 individuals with at least one primary care visit within the Sutter Health system in Northern California between 2014–2018. Based on MPA score, we initially identified 92,673 migraine cases and 278,019 matched controls. We excluded 3,065 cases and 7,327 controls who resided outside of 27 Northern California counties; 29 cases and 100 controls who lacked block group-level poverty data; and 4 cases and 28 controls missing $PM_{2.5}$ data (**Supplemental Figure 4.1**). The final study population included 89,575 cases and 270,564 controls in 27 counties (**Supplemental Figure 4.2, Supplemental Figure 4.3**).

Migraine cases were most common between the ages of 30–44 years (N = 33,036, 36.9%) and occurred predominantly among females (N = 73,908, 82.5%). Migraine cases were more likely to be non-Hispanic white as compared with controls (58.7% versus 48.2%) and had more frequent primary care outpatient encounters and outpatient neurologist visits (**Table 4.1**). The 2015 average annual concentrations of $PM_{2.5}$ and NO_2; methane emission rates; and the location of active oil and gas wells are depicted in **Figure 4.1**. The median $PM_{2.5}$ concentration at patient addresses was 8.7μg/m^3 (min = 3.7, max = 13.3) and the median NO_2 concentration was 7.7ppb (min = 1.1, max = 15.2). Of 564 high-emitters surveyed in the state, 60 (10.6%) were located within 10km of study participants, including 35 dairies/landfills and 25 other types of high-emitters.

Case-Control Analysis

In our case-control analysis we observed only linear associations between exposure and migraine case status. We found some evidence for an association between migraine case status and block-group level NO_2 concentration. We estimated that for every 5ppb increase in annual average NO_2 concentration the odds of migraine case status increased by 1.02 times (95% CI: 1.00, 1.05). We also estimated that for every 100,000 kg/hour increase in IDW sum of methane emissions within 10km, the odds of migraine case status also increased (OR = 1.04, 95% CI: 1.00, 1.08). We found no evidence of an association between migraine case status and block-group level $PM_{2.5}$ concentrations or for active oil or gas wells within 10km **(Figure 4.2, Supplemental Table 4.1A**). In our secondary analysis with dichotomized methane emission and active wells, we found no association between any high methane emitter or any active well within 10km and migraine **(Supplemental Table 4.1B)**.

Case-Case Analysis

In our case-case analysis, meant to evaluate the association between environmental exposures and migraine frequency/severity, we observed mostly linear relationships, except for the association between $PM_{2.5}$ and odds of any migraine ED visit during the study period (**Supplemental Figure 4.4**). For the other severity outcomes, we found that each 5μg/m^3 increase in annual average block-group level $PM_{2.5}$ concentration was associated with increased frequency of outpatient neurology visits (RR = 1.18, 95% CI: 1.09, 1.29), increased frequency of migraine-specific urgent care visits (RR = 3.09, 95% CI: 2.28, 4.18) and MPA score greater than 100 (OR = 1.14, 95% CI: 1.07, 1.22). We found no evidence of an association between increased $PM_{2.5}$ concentration and frequency of prescribed triptans (RR = 1.03, 95% CI: 0.97, 1.10). Increased block group-level NO_2 concentration was not associated with triptans, outpatient neurology visits or MPA score, but we found that each 5 ppb increase in NO_2 concentration was associated with increased frequency of migraine-specific urgent care visits (RR = 1.22, 95% CI: 1.02, 1.46) and with increased odds of having at least one migraine-specific ED visit during follow-up (OR = 1.16, 95% CI: 1.05, 1.29) (**Figure 4.3, Supplemental Table 4.2A**).

A 100,000-unit increase in the IDW sum of overall methane emissions within 10km was associated with increased frequency of migraine-specific urgent care visits (RR = 1.12, 95% CI: 0.92, 1.36). Having any methane emitter within 10km was also associated with increased frequency of urgent care visits (RR = 1.32, 95% CI: 1.14, 1.54) (**Figure 4.3, Supplemental Tables 4.2A and 4.2B**). Proximity to high-emitters was not associated with the frequency of triptan prescriptions, outpatient neurologist visits, migraine-specific ED visits, or MPA score. Presence of any active oil and gas wells within 10km was associated with increased frequency of outpatient neurologist visits (RR = 1.09, 95% CI: 1.03, 1.16), frequency of migraine-specific urgent care visits (RR = 1.43, 95% CI: 1.21, 1.70), and odds of at least one migraine-specific ED encounter per person-year of follow-up (OR = 1.11, 95% CI: 1.00, 1.24). We found no evidence of an association between our continuous measure of active oil and gas wells and any of the five measures of migraine severity (**Figure 4.3, Supplemental Tables 4.2A and 4.2B**).

<u>Sensitivity Analyses</u>
We conducted a sensitivity analysis in which we separately considered dairies and landfills versus all other high methane emitters. Overall, these findings were largely consistent with our main findings for both the case-control and case-case analyses; the association was stronger for dairies and landfills (RR = 1.18, 95% CI: 0.37, 3.87) than for other high-emitters (1.08, 95% CI: 0.85, 1.36), albeit with widely overlapping confidence intervals. In re-analysis of the case-control and case-cases studies with additional controls for BMI category and marital status, results did not differ from those of our primary analysis (**Supplemental Figures 4.5 and 4.6**). Results were also unchanged when we incorporated distance to the nearest Sutter hospital in the ED visit case-case analyses (**Supplemental Figure 4.7**).

4.5 Discussion

Past research links short-term exposure to a range of air pollutants with ED visits migraine headache. Our study builds upon previous studies and considers the implications of long-term environmental exposures for migraine. Using data from the Sutter Health EHR database in Northern California, we examined relationships between a wide range of environmental exposures–including $PM_{2.5}$, NO_2, high methane emitters, and oil and gas wells–and both migraine headache and headache severity among patients with migraine. Our case-control analysis revealed increased odds of exposure to NO_2 and high methane emitters among patients with migraine as compared with frequency-matched population controls without clinical diagnosis of migraine. In our case-case analysis, migraine severity–as measured by frequency of triptan prescriptions, outpatient neurology visits, migraine-specific urgent care and ED visits, and MPA score–was most strongly and consistently associated with average $PM_{2.5}$ and NO_2 exposure.

Research to date has focused primarily on short-term exposure to air pollutants as a trigger for migraine. Although relatively few studies have focused on chronic exposure, evidence to date nevertheless suggests that chronic exposure to common pollutants may be important in the etiology, severity, or frequency of headache including migraine. Using linked records from the Taiwan National Health Insurance Research Database and Taiwan Air Quality Monitoring Database, Hong et al. (2020) found that frequency of recurrent headaches among children younger than 18 years of age increased with higher-level exposure to several air pollutants including $PM_{2.5}$, CH_4, NO_2, and total hydrocarbons (Hong et al. 2020). Adetona et al. (2020) conducted a cross-sectional study among residents of a community adjacent to a large open landfill in Lagos, Nigeria. Results of that study indicated that chronic exposure to emissions from open combustion of municipal solid waste—a major source of particulate matter, polycyclic aromatic hydrocarbons, and toxicants such as polychlorinated biphenyls and brominated flame retardants—was associated with increased odds of daily occurrence of headache (Adetona et al. 2020). Moreover, in animal models, chronic exposure to acrolein, which is prevalent in both indoor and outdoor air pollution, yielded physiologic changes consistent with migraine (Kunkler et al. 2015, 2018).

Results of the present study further demonstrate the potential importance of long-term residential exposures for migraine severity. One important implication of these findings is that in more heavily polluted communities, individuals may be more likely to suffer from migraines or may suffer from more frequent headaches. The existing literature consistently demonstrates the disproportionate burden of air pollution in already disadvantaged communities (Colmer et al. 2020; O'Neill et al. 2003; Woo et al. 2019), and the substantial economic and social costs associated with migraine in the United States (De Lissovoy and Lazarus 1994; Ferrari 1998; Hu et al. 1999; Lipton et al. 2001b, 2007). Our findings therefore motivate careful examination of the extent to which disparate levels of exposure to harmful emissions and levels of community air pollution translate to greater burden of migraine headache and the associated economic and social costs particularly in already disadvantaged communities.

To our knowledge, ours is the first study to examine the implications of exposure to high methane emitters for migraine; we identified an association between high methane emitter exposure and migraine case status but not migraine severity. High methane emitters included dairies and waste lagoons, landfills, power plants, refineries, wastewater treatment facilities, and oil and gas production and distribution infrastructure. Although methane itself is not directly toxic to humans, it is often co-emitted with other noxious compounds. The heterogeneous group of high-emitters considered in this study also produce a wide range of co-pollutants including volatile organic compounds, ammonia, hydrogen sulfide, and particulate matter, several of which are odorous (Casey et al. 2015b; Garcia-Gonzales et al. 2019b; Staines 2004). Methane also contributes to the formation of ground-level ozone, previously implicated as a trigger for migraine headache (Chen et al. 2015; Chiu et al. 2015). In addition, high-emitters, such as oil and gas wells, produce noise pollution (Hays et al. 2017b). Both noise and odors have been consistently linked with migraine headache (Charles 2013; Eross et al. 2007; Prince et al. 2004; Wöber et al. 2006).

Importantly, we assigned high methane emitter exposure based on data collected between 2016–2018, while we included migraine cases in the Sutter EHR database between 2014–2018. This complicates the temporal ordering of exposure and response. However, reverse causality seems an implausible alternative explanation for our results, as we know of no reason that individuals with migraine would cause systematic increases in local high methane emitter exposure or would move closer to a high methane emitter post-diagnosis. It is possible, however, that our findings reflect residential sorting of individuals predisposed to migraine into localities where methane emissions are higher on average (Spielman et al. 2013; Watson 2009). In the U.S., migraine follows a social gradient and is more common among lower-income individuals who are also more likely to live in more polluted neighborhoods (Lipton et al. 2001a, 2007). We aimed to address this important source of confounding by adjusting for patient Medicaid use and block-group-level population density and poverty. Future research should specifically examine co-pollutants that may explain the apparent link between methane emissions and migraine, and to disentangle the role of residential sorting and confounding by socioeconomic status from any etiologic role that methane plays in the onset or exacerbation of migraine headaches.

Unlike several prior studies that rely on ED visits as a rough proxy for disease severity, our case-case analysis considered a more comprehensive set of proxies obtained from EHR data including non-emergency migraine-specific healthcare visits, migraine-related medication use, a validated migraine severity score, and overall neurology visits among patients with migraine. We also used splines to consider potential non-linearities in exposure-response relationships between each environmental exposure and our migraine severity outcome measures. Consistent with past research (Lipton et al. 2001a, 2007), we observed an association between NO_2 exposure and migraine severity as measured by migraine-specific urgent care visits and migraine-specific ED visits even at NO_2 levels well below the current national standards (our population-average annual exposure was around 8 ppb compared to the U.S. Environmental Protection Agency annual standard of 53 ppb).

Past research finds an association between short-term exposure to $PM_{2.5}$ and migraine-specific ED visits (Lipton et al. 2001a, 2007). Our analysis demonstrated an association with long-term, annual average $PM_{2.5}$ across a more comprehensive set of clinical proxies for headache severity, including outpatient neurology visits and migraine-specific urgent care visits. For ED visits, we found a paradoxical inverse u-shaped exposure-response wherein individuals with the lowest and highest levels of average $PM_{2.5}$ had the lowest odds of ED visit. This relationship persisted even after we incorporated additional statistical controls for distance to nearest Sutter ED. As our analysis differs from previous studies that consider short-term $PM_{2.5}$ levels and risk of ED visits, this finding could reflect misalignment of the examined exposure window (annual average $PM_{2.5}$) with an acute outcome (ED visits).

Communities with higher annual $PM_{2.5}$ concentrations may also have higher peak and long-term average exposure that gives rise to ED visits. We know of no research that demonstrates higher levels of $PM_{2.5}$ as protective against migraine headaches. This relationship could reflect residential sorting where individuals with migraine move out of high $PM_{2.5}$ communities. As migraine-specific emergency department visits are relatively rare in these data, we suspect that the observed relationship is driven by relatively less frequent use of emergency departments for headache among individuals living in the few counties with the highest $PM_{2.5}$ levels. This finding also implies possible geographic disparities in either access to or use of care for severe migraine headaches unrelated to proximity or insurance status that should be explored in future research.

The association between $PM_{2.5}$ and migraine severity may be partly explained by correlation between $PM_{2.5}$ and other exposures known to precipitate migraine headache (namely, noise and noxious odors) (Charles 2013; Eross et al. 2007; Prince et al. 2004; Wöber et al. 2006). $PM_{2.5}$, is known to activate the sympathetic nervous system, result in systemic inflammation, and trigger cardiovascular events (Feng et al. 2016; Pope et al. 2004), and may also directly result in migraine. The smallest fraction of the $PM_{2.5}$ particles, ultrafine particulate matter ($\leq 0.1\mu m$ in diameter (Davidson et al. 2005)), may have a disproportionately large role. Ultrafine particles–unlike the larger component particles of $PM_{2.5}$–can transverse the blood-brain barrier and reach the brain directly through the olfactory bulb (Schraufnagel 2020).

Despite making up just a small portion of the total $PM_{2.5}$ mass concentration, these circumstances raise the possibility that the apparent association between $PM_{2.5}$ and migraine severity in this and previous studies could be partially explained by neurotoxic effects secondary to exposure to the ultrafine component of $PM_{2.5}$ (Costa et al. 2017; Win-Shwe and Fujimaki 2011). The U.S. EPA does not regulate ultrafine particulate matter, meaning exposure estimates are sparse and epidemiologic studies rare. Future migraine research should aim to evaluate the effects of ultrafine particles on migraine and disentangle the effects of concomitant exposure to noise, odor, $PM_{2.5}$, and ultrafine particles.

Limitations

Our analyses include all individuals with migraine followed from 2014-2018 but do not distinguish between individuals with previously diagnosed migraine at the beginning of the study period (i.e., prevalent cases) and individuals diagnosed with migraine throughout the study period (i.e., incident cases). This makes ascertainment of an etiologic role of environmental exposures in either migraine onset or exacerbation challenging. As discussed previously, we cannot eliminate the possibility that our findings may reflect residential sorting, wherein individuals with existing migraine are more likely to reside in health-harming communities, for example those of lower socioeconomic status or with higher levels of pollutants. Alternatively, individuals with migraine and the financial means to do so may choose to leave communities with environmental exposures that trigger their headaches. The direction and magnitude of bias attributable to residential sorting is therefore difficult to anticipate.

Although our analyses include individuals with migraine followed from 2014–2018, exposures were either measured at the beginning of the study period in 2015 (annual average $PM_{2.5}$, NO_2, and presence of oil and gas wells) or as values between 2016–2018 (high methane emitter emissions and presence). We assume relatively stable levels of long-term air pollution and oil and gas well exposure during the study period. High methane emitter measurements took place between 2016–2018, but emission trends likely vary over time. Exposures were also assigned based on a single residential address on the index date and therefore do not capture exposure accrued during time spent outside the home and also do not reflect potential moves between 2014–2018. Future research should endeavor to incorporate time-varying measures of air pollution, oil and gas wells, and methane emissions in relation to migraine onset and exacerbation in order to better characterize the dynamic relationship between environment and migraine.

Residential addresses were ascertained in October 2019 after the study period. Selection bias could result, for example, if individuals with migraine headache in highly polluted counties moved to less polluted counties *outside* of the Sutter catchment areas. Because a small minority of individuals lived outside of the Sutter catchment area in October 2019 (3.3% of cases and 2.6% of controls), we expect any resultant bias to be minimal. Some differential exposure misclassification could also arise if individuals with migraine headache in highly polluted counties moved to less polluted counties *within* the Sutter catchment area, leading to systematic underestimation of long-term exposures among cases, and therefore, underestimation of effect estimates.

While our study incorporates a more comprehensive set of proxy measures for migraine headaches as compared with previous studies (which typically relied on migraine ED visits), we lacked any direct measure of headache frequency among patients with migraine (e.g., headache diaries). Our results rest on patients seeking clinical care for migraine. If individuals with higher levels of environmental exposure were systematically less likely to seek migraine treatment, our results may be attenuated. Headache diaries would circumvent this problem and further examination of the relationship between migraine and the environment in datasets where direct measures of headache frequency are available (Cooke et al. 2000; Giffin et al. 2003; Moloney et al. 2009) would further our understanding of this relationship.

Fourth, our analysis includes a comprehensive set of potential confounding variables. Nevertheless, we note the absence of several critical variables–including individual-level income, educational attainment, and employment status–that may be important confounders in studies that use treatment seeking as a proxy for headache severity, given past research showing that migraine plays a key role in disability, absence from work or school, and that migraine follows a social gradient and is less common in wealthier individuals (Gilligan et al. 2018; Lipton et al. 2007; Porter et al. 2019). Further, we lacked information on environmental noise pollution, which may trigger migraines (Borkum 2016) and often co-occurs with sources of air pollution.

Finally, we drew participants from a single healthcare system in Northern California. This may limit generalizability to other populations including individuals who are uninsured or have limited health insurance. Northern California also differs meaningfully from the rest of the U.S. in the quality and extent of environmental exposures and population demographics. The relationship between migraine and environment may differ by region, season, and based on individual characteristics. This motivates ongoing study of the relationship between migraine in the environment in varied contexts.

4.6 Conclusions

In this study, we demonstrate an association between long-term NO_2 and high methane emitter exposure and odds of being a migraine patient. We also find annual average NO_2 and $PM_{2.5}$ exposure associated with migraine headache severity. Our study expanded the scope of environmental pollutants considered as risk factors for migraine and included numerous measures of migraine severity derived from EHR data and contributes to the existing literature on migraine and the environment by explicitly considering long-term exposure to common pollutants. These findings illustrate the potential role of ambient air pollution for prevalent migraine and migraine severity. Future studies are needed that establish the temporal ordering of exposure and outcome and the relevant exposure period as well as that determine the most relevant air pollutants. In addition, researchers should consider the potential heterogeneity in the relationship between migraine and the environment across different geographic contexts and within population subgroups. Such studies could identify environmental risk factors on which we could intervene to reduce the population burden of migraine.

4.7 Tables and Figures

Table 4.1. Patient demographics, healthcare utilization, and environmental exposures for migraine cases and controls from Sutter Health in Northern California, 2014–2018

	Migraine cases N = 89,575	**Controls**[a] N = 270,564
Patient Demographics		
Age Category, N (%)		
18 – 29 years	16952 (18.9)	51112 (18.9)
30 – 44 years	33036 (36.9)	99792 (36.9)
45 – 54 years	19226 (21.5)	58169 (21.5)
55 – 64 years	12578 (14.0)	38093 (14.1)
≥ 65 years	7783 (8.7)	23399 (8.7)
Sex, N (%)		
Female	73908 (82.5)	223230 (82.5)
Male	15667 (17.5)	47334 (17.5)
Race/Ethnicity, N (%)		
Non-Hispanic		
Asian	9278 (10.4)	52794 (19.9)
Black	3685 (4.1)	10253 (3.8)
White	52579 (58.7)	130418 (48.2)
Other	11351 (12.7)	41907 (15.5)
Hispanic	12682 (14.2)	34192 (12.6)
Marital Status, N (%)		
Divorced/Separated/Widowed	7444 (8.3)	18881 (7.0)
Married/Significant Other	51390 (57.4)	155644 (57.5)
Single	22659 (25.3)	63801 (23.6)
Other/Unknown	8082 (9.0)	32238 (11.9)
Body Mass Index Category (kg/m^3), N (%)		
Underweight (<18.5)	1672 (1.9)	5636 (2.0)
Normal (18.5-24.9)	33801 (37.7)	112014 (41.4)
Overweight (25-29.9)	26969 (30.1)	79209 (29.3)
Obese Class 1 (30-34.9)	14595 (16.3)	39405 (14.6)
Obese Class 2 (35-39.9)	6835 (7.6)	17388 (6.4)
Obese Class 3 (40+)	4614 (5.2)	11658 (4.3)
Missing	1089 (1.2)	5254 (1.9)
Block Group-Level Variables, Median (IQR)		
Percent Poverty	7.2 (3.5, 14.3)	6.6 (3.2, 13.1)
Population Density (individuals per km^2)	2211 (901, 3593)	2292 (954, 3592)
Medicaid Beneficiary, N (%)		
Yes	6929 (7.7)	15105 (5.6)
No	82646 (92.3)	255459 (94.4)
Healthcare Utilization		

Encounters per person-year		
Primary care, Median (IQR)	2.4 (1.4, 4.0)	1.9 (1.2, 3.1)
Neurology, Mean (SD)	1.2 (3.4)	0.2 (1.0)
Urgent Care (Migraine-Specific), Mean (SD)	0.2 (2.7)	--
Emergency (Migraine-Specific), N (%)	0.1 (0.7)	--
≥ 1 visit during the study period	3987 (4.5)	--
< 1 visit during the study period	85588 (95.5)	--
Triptan prescriptions per person-year, mean (SD)	0.6 (2.6)	
MPA Score – N (%)	66.6 (31.5)	--
≤ 10	59599 (66.5)	--
>10	29976 (33.5)	--
Environmental Exposures		
Air Pollutants, Median (IQR)		
NO_2, ppb	7.7 (5.7, 10.2)	8.1 (5.9, 10.4)
$PM_{2.5}$, μg/m^3	8.7 (7.8, 9.6)	8.9 (7.8, 9.7)
CH_4 Emissions		
Any high methane emitter within 10km, N (%)	18457 (20.6)	57224 (21.1)
Total IDW emissions in kg/hour, Mean (SD)	21,461 (192,973)	26,070 (180,548)
Active Oil and Gas wells		
Any oil or gas well within 10km, N (%)	13179 (14.7)	37010 (13.7)
Total IDW wells, Mean (SD)	604 (6468)	603 (6459)

IDW, inverse-distance weighted; IQR, interquartile range; MPA, migraine probability algorithm
[a] Frequency-matched on age category, sex, year of entry into Sutter primary care, and primary-care follow-up time (0-6 months, 7-24 months, ≥ 24 months).

Figure 4.1. Distribution of environmental exposures within study region. Block group level 2015 annual average concentration of A. $PM_{2.5}$ and B. NO_2. C. Methane emission rate based on the California Methane Survey, conducted between 2016–2018. D. Location of active oil and gas wells as of December 2015.

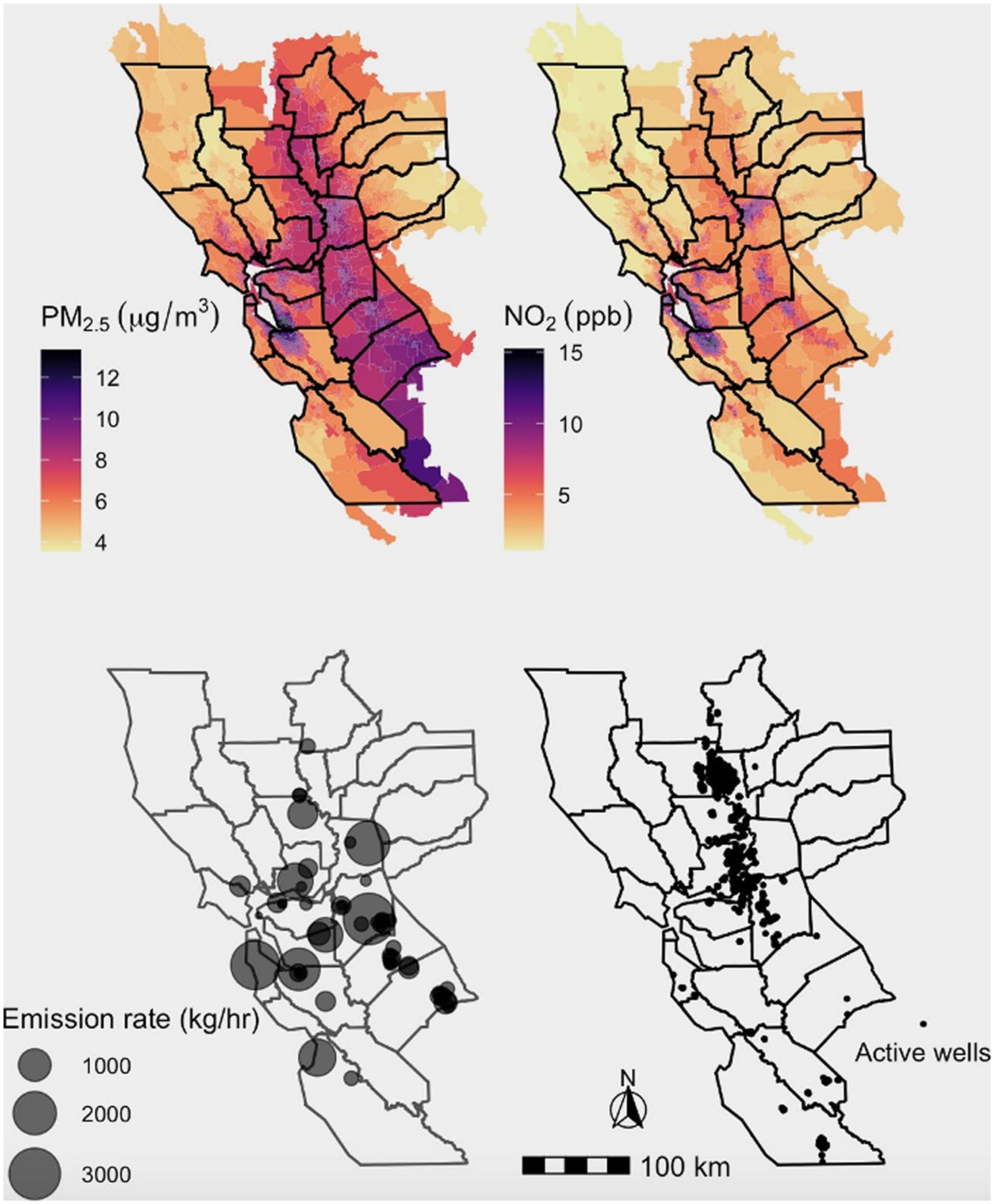


Figure 4.2. Association between environmental exposures and odds of being a migraine case versus control. Results from a mixed logistic model with a random intercept for county adjusted for individual-level age category (18-29, 30-44, 45-54, 55-64, ≥65), race/ethnicity (Hispanic, non-Hispanic Asian, non-Hispanic-Black, non-Hispanic White, and non-Hispanic other), sex, Medicaid use, number of primary care visits per person-year during the study period, and block group-level population density and poverty. OR are per 5μg/m^3 for $PM_{2.5}$, per 5ppb for NO_2, per 100,000 kg/hour increase in IDW sum of methane emissions within 10km for high-emitters, and per 1,000-unit increase in IDW sum of all wells within 10km for active oil and gas wells.

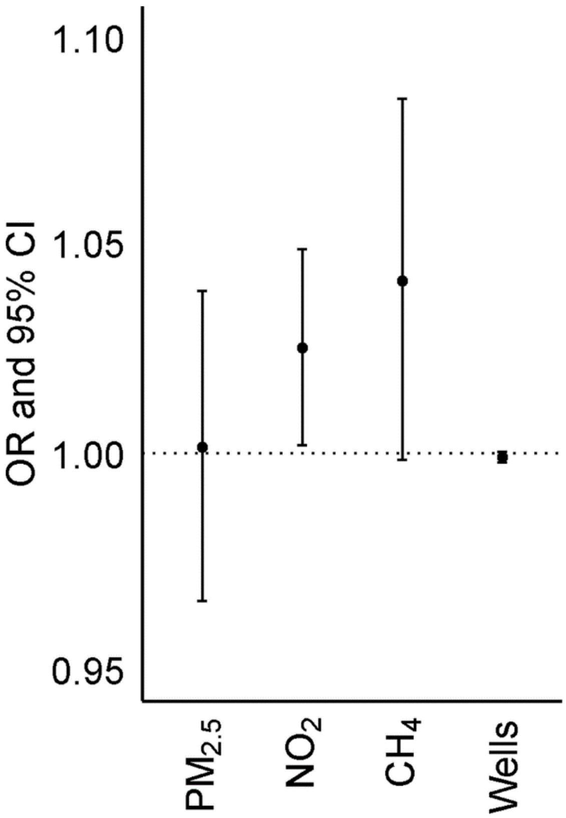
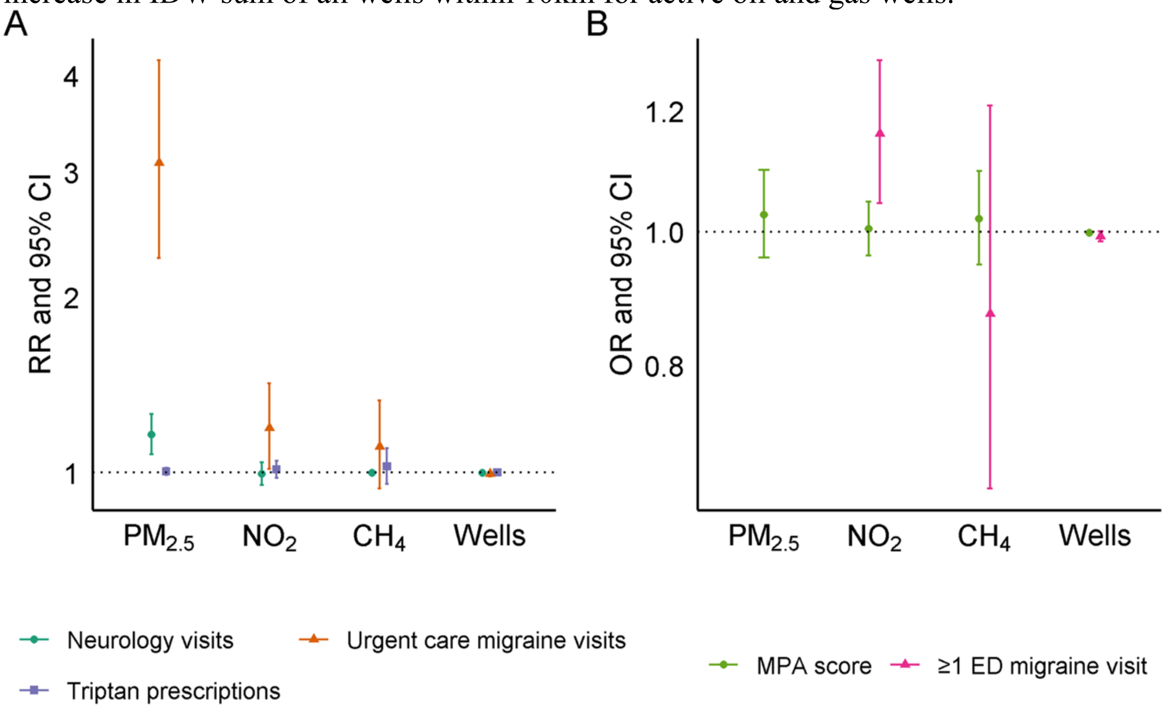


Figure 4.3. Association between environmental exposures and severity of migraine case status. Associations estimated with mixed logistic and negative binomial models with random intercepts for county adjusted for individual-level age category (18-29, 30-44, 45-54, 55-64, ≥65), race/ethnicity (Hispanic, non-Hispanic Asian, non-Hispanic-Black, non-Hispanic White, and non-Hispanic other), sex, Medicaid use, number of primary care visits per person-year during the study period, and block group-level population density and poverty. Neurology visits, urgent care migraine-specific visits, and triptan prescriptions were parameterized as continuous counts per person-year and analyzed using negative binomial models **(Panel A)**. ED migraine visits were dichotomized as zero vs. ≥ 1 during the study period, and MPA score as >100 versus less **(Panel B)**. ORs and RRs are per 5μg/m^3 for $PM_{2.5}$, per 5ppb for NO_2, per 100,000 kg/hour increase in IDW sum of methane emissions within 10km for high-emitters, and per 1,000-unit increase in IDW sum of all wells within 10km for active oil and gas wells.



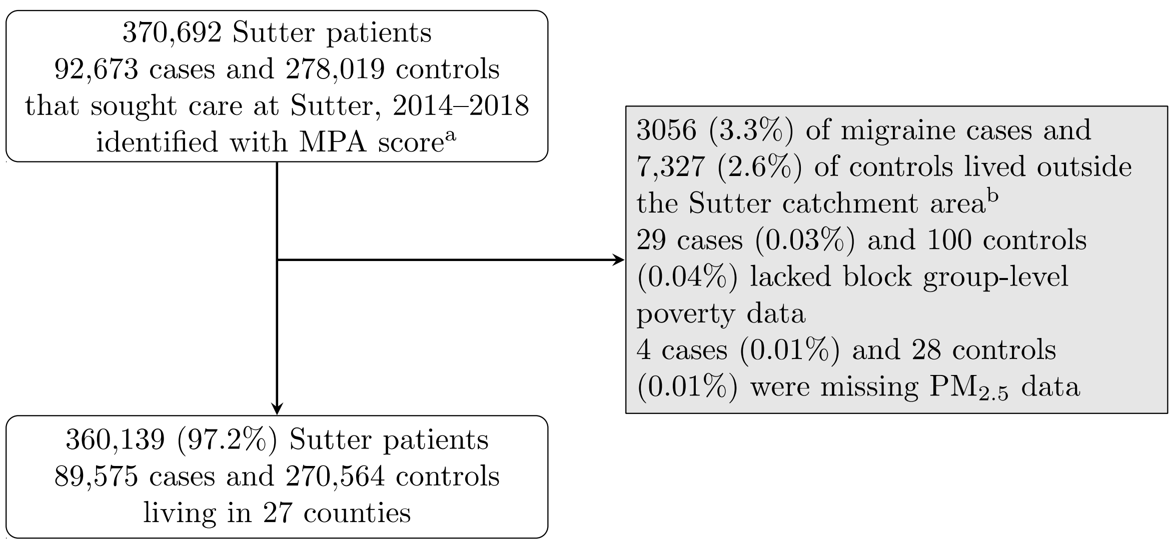
4.8 Declarations

Ethics approval and consent to participate: The Columbia University (Protocol #: AAAT0085), University of California, Berkeley (Protocol #: 2013-10-5693), and Sutter Health (IRBNet #: 1452543-1) Institutional Review Boards approved this study.

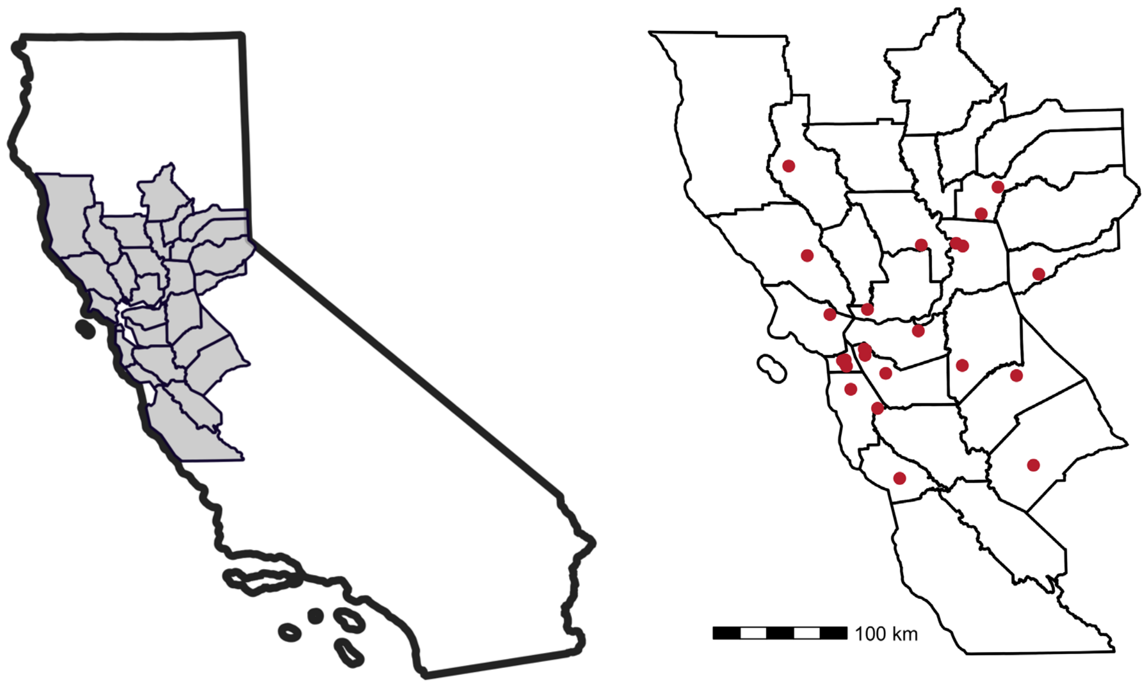
Availability of supporting data: The Sutter Health electronic health record data are considered Protected Health Information under the Health Insurance Portability and Accountability Act of 1996 (HIPAA) in the United States, and as such are not publicly available. $PM_{2.5}$ and NO_2 data are available for download at: https://www.caces.us/data. Methane data are available via https://www.nature.com/articles/s41586-019-1720-3#data-availability. Oil and gas well data are available at https://www.conservation.ca.gov/calgem/Pages/Oil-and-Gas.aspx.

4.9 Supplemental Information Chapter 4. Air pollution, high methane emitters, and oil and gas wells in Northern California: the relationship with migraine headache prevalence and exacerbation

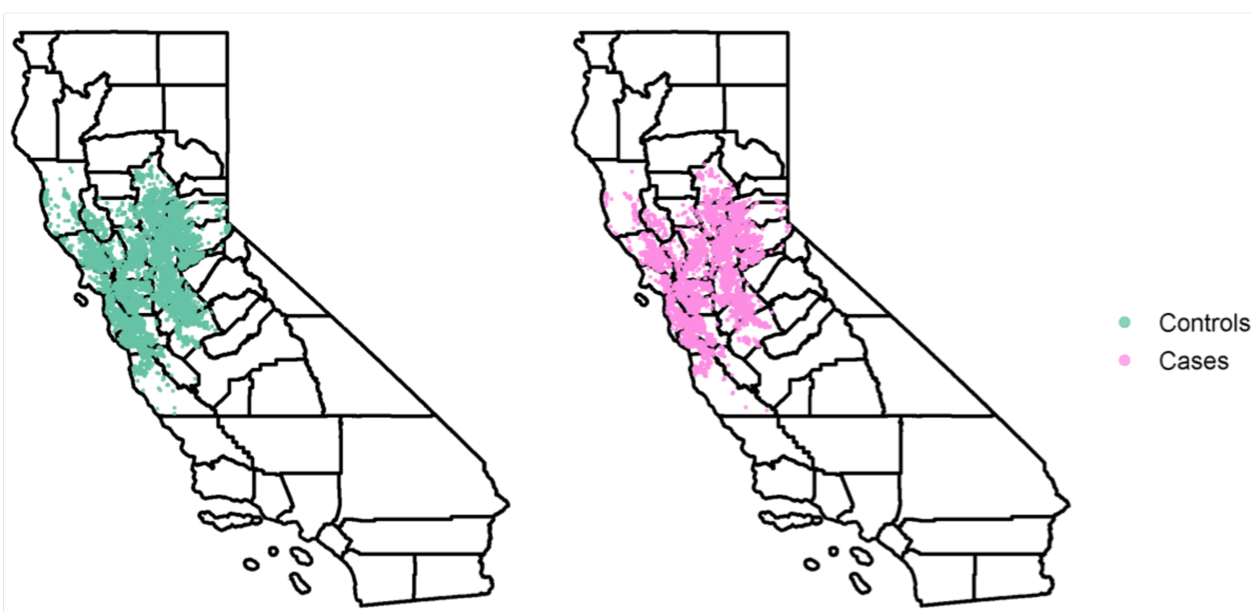
Supplemental Figure 4.1. Ascertainment of migraine cases and controls from Sutter Health electronic health record data, 2015–2018.
[a] The migraine probability algorithm (MPA) is based on migraine-related *International Classification of Diseases*-9 and 10 (*ICD*-9 and *ICD*-10) codes in the primary or secondary position in the outpatient or emergency department setting, on the patient's Significant Health Problem List, migraine prescription medications, and outpatient *ICD* codes related to cluster headache.
[b] Catchment counties include Alameda County, Amador County, Butte County, Colusa County, Contra Costa County, El Dorado County, Lake County, Mendocino County, Monterey County, Napa County, Marin County, Merced County, Nevada County, Placer County, Sacramento County, San Benito County, San Francisco County, San Joaquin County, San Mateo County, Santa Clara County, Santa Cruz County, Solano County, Sonoma County, Stanislaus County, Sutter County, Yolo County, Yuba County.



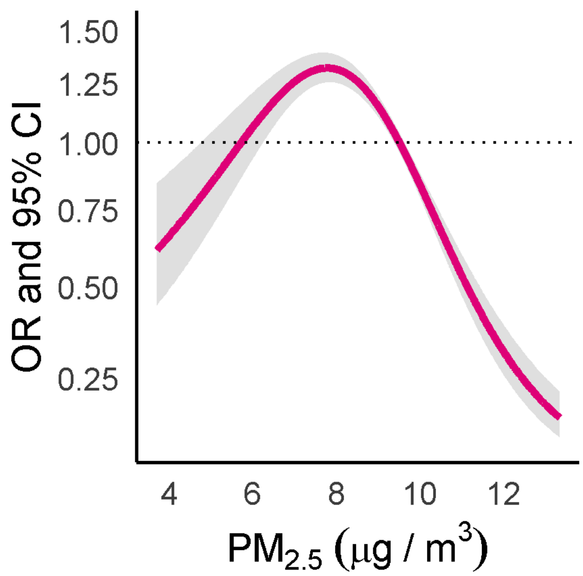
Supplemental Figure 4.2. Counties included in the analysis in Northern California (left) and distribution of Sutter hospitals (right).



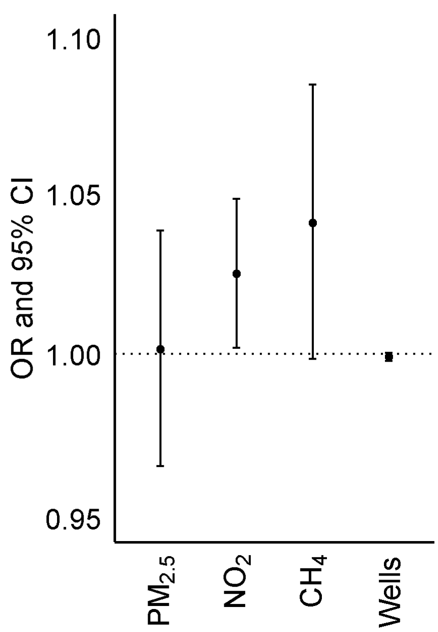
Supplemental Figure 4.3. Distribution of migraine cases and controls.



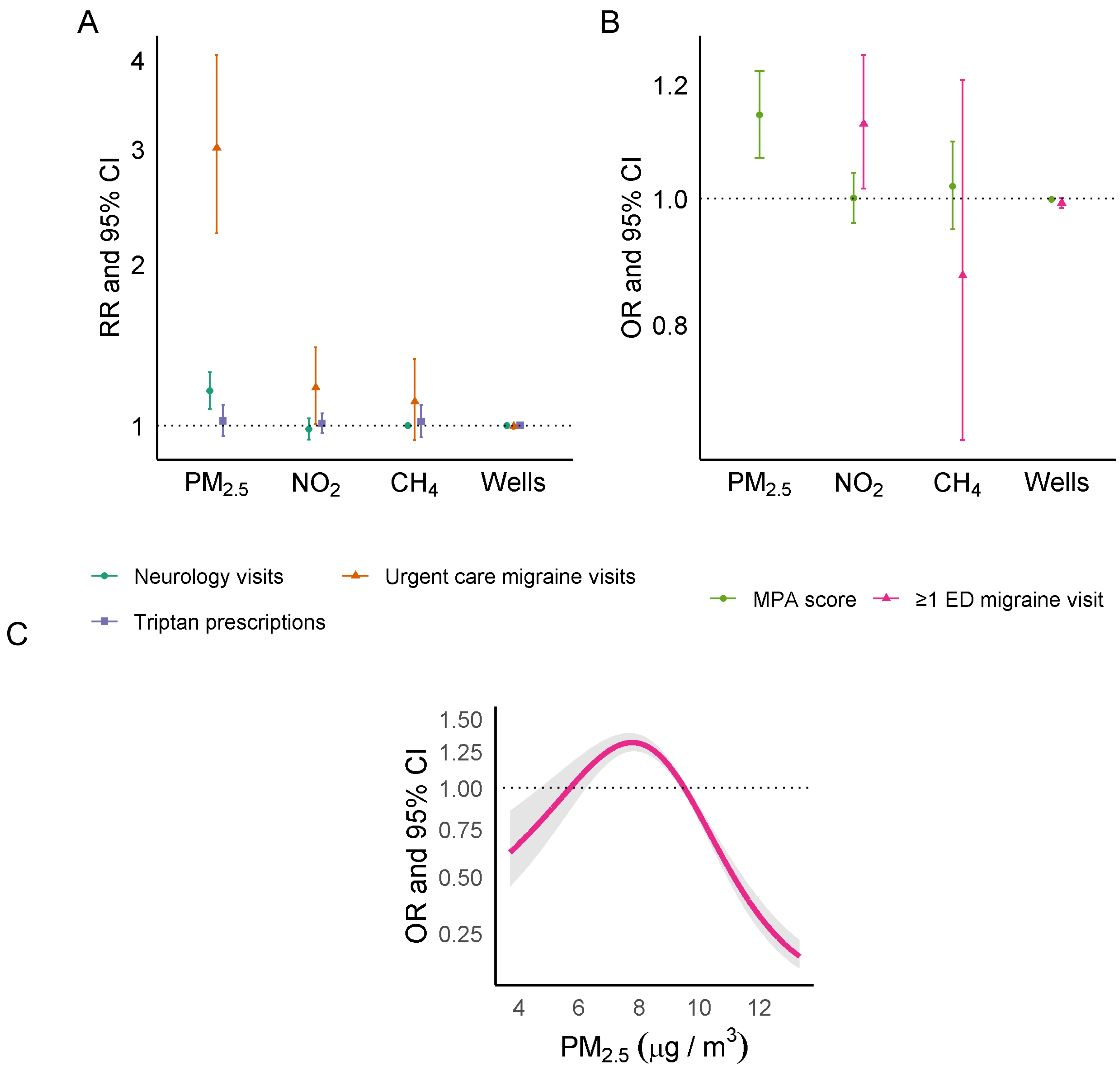
Supplemental Figure 4.4. Flexible dose-response between levels of $PM_{2.5}$ ($\mu g/m^3$) and odds of having ≥ 1 ED visit over the course of the study period. From mixed logistic models with penalized smoothing splines for $PM_{2.5}$, random intercept for county, adjusted for individual-level age category (18-29, 30-44, 45-54, 55-64, ≥ 65), race/ethnicity (Hispanic, non-Hispanic Asian, non-Hispanic-Black, non-Hispanic White, and non-Hispanic other), sex, Medicaid use, number of primary care visits per person-year during the study period, and block group-level population density and poverty.



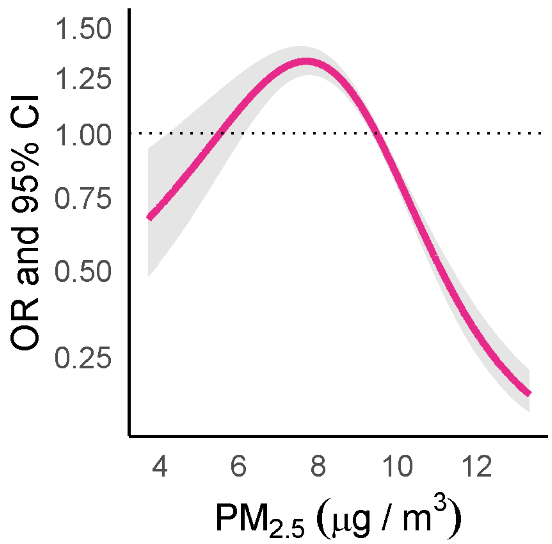
Supplemental Figure 4.5. Association between environmental exposures and odds of being a migraine case versus control. Results from a mixed logistic model with a random intercept for county adjusted for BMI category (underweight < 18.5; normal weight 18.5 – 24.9; overweight 25 – 29.9; obese class I 30 – 34.9; obese class 2 30 – 34.9; obese class 3 40+; missing), marital status (divorced, separated widowed; married or significant other; single; other or unknown), individual-level age category (18-29, 30-44, 45-54, 55-64, ≥65), race/ethnicity (Hispanic, non-Hispanic Asian, non-Hispanic-Black, non-Hispanic White, and non-Hispanic other), sex, Medicaid use, number of primary care visits per person-year during the study period, and block group-level population density and poverty. OR are per 5μg/m^3 for $PM_{2.5}$, per 5ppb for NO_2, per 100,000 kg/hour increase in IDW sum of methane emissions within 10km for high-emitters, and per 1,000-unit increase in IDW sum of all wells within 10km for active oil and gas wells.



Supplemental Figure 4.6. Association between environmental exposures and severity of migraine case status. Associations estimated with mixed logistic and negative binomial models with random intercepts for county adjusted for BMI category (underweight < 18.5; normal weight 18.5 – 24.9; overweight 25 – 29.9; obese class I 30 – 34.9; obese class 2 30 – 34.9; obese class 3 40+; missing), marital status (divorced, separated widowed; married or significant other; single; other or unknown), individual-level age category (18-29, 30-44, 45-54, 55-64, ≥65), race/ethnicity (Hispanic, non-Hispanic Asian, non-Hispanic-Black, non-Hispanic White, and non-Hispanic other), sex, Medicaid use, number of primary care visits per person-year during the study period, and block group-level population density and poverty. Neurology visits, urgent care migraine-specific visits, and triptan prescriptions were parameterized as continuous counts per person-year and analyzed using negative binomial models (**Panel A**). ED migraine visits were dichotomized as zero versus ≥ 1 during the study period, and MPA score as >100 versus less. ORs and RRs are per 5μg/m^3 for $PM_{2.5}$, per 5ppb for NO_2, per 100,000 kg/hour increase in IDW sum of methane emissions within 10km for high-emitters, and per 1,000-unit increase in IDW sum of all wells within 10km for active oil and gas wells (**Panel B**). Non-linear exposure response curve for the association between levels of $PM_{2.5}$ (μg/m^3) and odds of having ≥ 1 ED visit over the course of the study period was modeled using mixed logistic models with penalized smoothing splines for $PM_{2.5}$ (**Panel C**).



Supplemental Figure 4.7. Association between $PM_{2.5}$ and migraine-specific ED visits, adjusted for distance to nearest Sutter hospital. Association estimated with a mixed logistic models with penalized smoothing splines for $PM_{2.5}$ and a random intercept for county adjusted for individual-level age category (18-29, 30-44, 45-54, 55-64, ≥65), race/ethnicity (Hispanic, non-Hispanic Asian, non-Hispanic-Black, non-Hispanic White, and non-Hispanic other), sex, Medicaid use, number of primary care visits per person-year during the study period, distance to nearest Sutter hospital in kilometers, and block group-level population density and poverty.



Supplemental Table 4.1A. Associations between continuous environmental exposures and migraine status

	Odds Ratio (95% CI)[a]
$PM_{2.5}$[b]	1.00 (0.97, 1.04)
NO_2[c]	1.02 (1.00, 1.05)
High methane emitters[d]	1.04 (1.00, 1.08)
Overall	1.04 (1.00, 1.08)
Dairies and Landfills[e]	1.07 (0.83, 1.39)
Other High-emitters[f]	1.05 (1.00, 1.10)
Active Oil & Gas Wells[g]	0.99 (0.99, 1.00)

a. From a mixed logistic model with a random intercept for county, adjusted for individual-level age category (18-29, 30-44, 45-54, 55-64, 65 or older), race/ethnicity (Hispanic, non-Hispanic Asian, non-Hispanic-Black, non-Hispanic White, and non-Hispanic other), sex, Medicaid use, number of primary care visits per person-year during the study period, and block group-level population density and poverty.
b. OR corresponds to a 5μg/m^3 increase in levels $PM_{2.5}$.
c. OR corresponds to a 5ppb increase in levels of NO_2.
d. OR corresponds to a 100,000 kg/hr increase in IDW sum of methane emissions within 10km.
e. Includes dairy/livestock manure, landfills, compost.
f. Includes powerplants, refineries, wastewater treatment facilities, oil and gas distribution (e.g., oil/gas compressors).
g. OR corresponds to a 1,000-unit increase in IDW sum of wells within 10km.

Supplemental Table 4.1B. Associations between dichotomized environmental exposures and migraine status

	Odds Ratio (95% CI)[a]
High methane emitters[b]	1.01 (0.99, 1.04)
Active Oil & Gas Wells[c]	1.01 (0.98, 1.04)

a. From a mixed logistic model with a random intercept for county, adjusted for individual-level age category (18-29, 30-44, 45-54, 55-64, ≥ 65), race/ethnicity (Hispanic, non-Hispanic Asian, non-Hispanic-Black, non-Hispanic White, and non-Hispanic other), sex, Medicaid use, number of primary care visits per person-year during the study period, and block group-level population density and poverty.
b. OR compares any high methane emitter within 10km versus none.
c. OR compares any active oil & gas wells within 10km versus none.

Supplemental Table 4.2A. Associations between continuous environmental exposures and measures of migraine severity

	Measures of Migraine Severity				
	Triptans[a] RR (95% CI)	**Neurology Visit[a]** RR (95% CI)	**Urgent Care Visit[a]** RR (95% CI)	**ED Visits[b]** OR (95% CI)	**MPA Score[b]** OR (95% CI)
$PM_{2.5}$[c]	1.01 (0.99, 1.02)	1.18 (1.09, 1.29)	3.09 (2.28, 4.18)	Non-Linear	1.14 (1.07, 1.22)
NO_2[d]	1.01 (0.98, 1.06)	0.99 (0.94, 1.05)	1.22 (1.02, 1.46)	1.16 (1.05, 1.29)	1.00 (0.96, 1.05)
High methane emitter[e]	1.03 (0.95, 1.12)	0.95 (0.85, 1.05)	1.12 (0.92, 1.36)	0.88 (0.63, 1.21)	1.01 (0.94, 1.09)
Overall	1.03 (0.95, 1.12)	0.95 (0.85, 1.05)	1.12 (0.92, 1.36)	0.88 (0.63, 1.21)	1.01 (0.94, 1.09)
Dairies and Landfills[f]	0.25 (0.02, 2.67)	0.91 (0.48, 1.71)	1.18 (0.36, 3.87)	0.96 (0.35, 2.68)	1.03 (0.67, 1.60)
Other High-emitters[g]	1.02 (0.93, 1.12)	0.91 (0.79, 1.05)	1.08 (0.85, 1.36)	0.94 (0.74, 1.24)	1.03 (0.95, 1.12)
Active Oil & Gas Wells[h]	1.00 (1.00, 1.01)	1.00 (1.00, 1.00)	0.99 (0.98, 1.01)	0.99 (0.98, 1.01)	0.99 (0.99, 1.00)

a. From mixed negative binomial models for frequency of triptans, neurology visits, and migraine-specific urgent care visits. All models included a random intercept for county, adjusted for individual-level age category (18-29, 30-44, 45-54, 55-64, ≥ 65), race/ethnicity (Hispanic, non-Hispanic Asian, non-Hispanic-Black, non-Hispanic White, and non-Hispanic other), sex, Medicaid use, number of primary care visits per person-year during the study period, and block group-level population density and poverty.

b. From mixed logistic models for ≥ 1 ED migraine visit during the study period and MPA score > 100. All models included a random intercept for county, adjusted for individual-level age category (18-29, 30-44, 45-54, 55-64, ≥ 65), race/ethnicity (Hispanic, non-Hispanic Asian, non-Hispanic-Black, non-Hispanic White, and non-Hispanic other), sex, Medicaid use, number of primary care visits per person-year during the study period, and block group-level population density and poverty.

c. Coefficients corresponds to a $5 \mu g/m^3$ increase in levels $PM_{2.5}$.

d. Coefficients corresponds to a 5ppb increase in levels of NO_2.

e. Coefficients corresponds to a 100,000 kg/hr increase in IDW sum of CH_4 emissions within 10km.

f. Includes dairy/livestock manure, landfills, and compost.

g. Includes powerplants, refineries, wastewater treatment facilities, oil and gas distribution (e.g., oil/gas compressors, gas distribution lines), oil and gas production (e.g., oil/gas waste lagoons, oil/gas plugged wells.

h. Coefficients corresponds to 1,000-unit increase in IDW sum of wells within 10km.

Supplemental Table 4.2B. Associations between binary environmental exposures and measures of migraine severity

	Measures of Migraine Severity				
	Triptans[a] RR (95% CI)	**Neurology Visit**[a] RR (95% CI)	**Urgent Care Visit**[a] RR (95% CI)	**ED Visits**[b] OR (95% CI)	**MPA Score**[b] OR (95% CI)
High methane emitting[c]	0.97 (0.91, 1.01)	0.99 (0.94, 1.04)	1.32 (1.14, 1.54)	1.06 (0.97, 1.16)	0.99 (0.95, 1.03)
Active Oil & Gas Wells[d]	0.99 (0.94, 1.04)	1.09 (1.03, 1.16)	1.43 (1.21, 1.70)	1.11 (1.00, 1.24)	1.02 (0.97, 1.07)

a. From mixed negative binomial models for frequency of triptans, neurology visits, and migraine-specific urgent care visits. All models included a random intercept for county, adjusted for individual-level age category (18-29, 30-44, 45-54, 55-64, ≥ 65), race/ethnicity (Hispanic, non-Hispanic Asian, non-Hispanic-Black, non-Hispanic White, and non-Hispanic other), sex, Medicaid use, number of primary care visits per person-year during the study period, and block group-level population density and poverty.

b. From mixed logistic models for ≥ 1 ED migraine visit during the study period and MPA score > 100. All models included a random intercept for county, adjusted for individual-level age category (18-29, 30-44, 45-54, 55-64, ≥ 65), race/ethnicity (Hispanic, non-Hispanic Asian, non-Hispanic-Black, non-Hispanic White, and non-Hispanic other), sex, Medicaid use, number of primary care visits per person-year during the study period, and block group-level population density and poverty.

c. Coefficient compares any high methane emitter within 10km versus none.

d. Coefficient compares any active oil & gas wells within 10km versus none.

Chapter 5: Climate justice and California's high methane emitters: An environmental equity assessment of community proximity and exposure intensity

Forthcoming paper**:** Casey JA, Cushing LJ, Depsky N, Morello-Frosch R (2021) Climate justice and California's methane super-emitters: An environmental equity assessment of community proximity and exposure intensity. *Environmental Science and Technology*.

5.1 Abstract

High methane emitters emit non-methane co-pollutants that are harmful to human health. Yet no prior studies have assessed disparities in exposure to high methane emitters with respect to race/ethnicity, socioeconomic status, and civic engagement. To do so, we obtained location, category (e.g., landfill, refinery), and emissions rate of California high methane emitters from Next Generation Airborne Visible/Infrared Imaging Spectrometer (AVIRIS-NG) flights conducted between 2016–2018. We identified block groups within 2km of high-emitters (exposed) and 5-10km away (unexposed) using dasymetric mapping and assigned level of exposure among block groups within 2km (measured via number of high methane emitter categories and total methane emissions). Analyses included 483 high-emitters. The majority were dairy/manure (n = 213) and oil/gas production sites (n = 127). Results from fully adjusted logistic mixed models indicate environmental injustice in locations of high methane emitters. For example, for every 10% increase in non-Hispanic Black residents, the odds of exposure increased by 10% (95% CI: 1.04, 1.17). We observed similar disparities for Hispanics and Native Americans, but not with indicators of socioeconomic status. Among block groups located within 2km, increasing proportions of non-White populations and lower voter turnout were associated with higher methane emitter emission intensity. Previously unrecognized racial/ethnic disparities in exposure to California high methane emitters should be considered in policies to tackle methane emissions.

5.2 Introduction

Since studies first documented the disproportionate siting of solid and hazardous waste facilities in Black communities in the 1980s, (Bullard 1983; Chavis and Lee 1987) subsequent environmental justice scholarship has demonstrated a consistent correlation between race, poverty, and pollution burden across diverse environmental hazards and geographies. Literature reviews conclude that people of color reside in neighborhoods with worse air quality(Clark et al. 2014; Grineski et al. 2017; Morello-Frosch and Jesdale 2006) and more environmental hazards (Brulle and Pellow 2006; Bullard et al. 2007; Mohai et al. 2009; Ringquist 2005) than White people in the United States. In California, environmental hazards including clean-up, hazardous waste, and solid waste sites are more regressively distributed with respect to race/ethnicity than poverty, suggesting that structural racism as opposed to class predominates in shaping inequalities (Cushing et al. 2015). This pattern is consistent with the history of legal racial discrimination in civil rights, housing, employment, and education that has produced staggering gaps in present day distributions of wealth across racial groups and led to persistent racial residential segregation (Brown 2016; Morello-Frosch 2002; Rothstein 2017).

In the current analysis, we investigate the social characteristics of communities near high methane emitter to assess potential environmental justice concerns. High methane emitters are point sources of large methane releases that span a wide range of industries. Though methane spends less time in the atmosphere than carbon dioxide (CO_2), its higher potency as a greenhouse gas makes its per-ton 'Global Warming Potential' some 84-86 times that of CO_2 over a 20-year period (Myhre et al. 2013). Compared to CO_2, therefore, reductions in methane emissions can more rapidly slow climate change. As a result, emissions reductions at large point sources of methane – including landfills, the oil and gas supply chain, livestock operations, and power plants – are being prioritized for near-term climate mitigation (Jackson 2009; Thurmond 2016). Atmospheric methane concentrations, however, have increased rapidly since 2008, driven primarily by the agriculture, waste, and fossil fuel sectors (Jackson et al. 2020). Moreover, studies suggest methane emissions in the U.S. substantially eclipse emissions inventories estimates, implying that methane releases are under-reported (Alvarez et al. 2018; Howarth 2019). In the natural gas sector, studies show that a small fraction of "high-emitters" (responsible for ~5% of leaks) contribute a disproportionate and under-reported amount of total methane emissions (~50% of emissions from leaks), usually due to abnormal and avoidable operating conditions, including equipment malfunctions (Brandt et al. 2016; Zavala-Araiza et al. 2015).

While high methane emitters are of significant interest due to their climate impacts, and specific types of high-emitters have been investigated from an environmental justice perspective (e.g., landfills, oil and gas wells, and concentrated animal feeding operations [CAFOs]), the possibility that high-emitters as a whole are disproportionately located in communities of color has not been examined. While not directly toxic to humans, methane is co-emitted with other pollutants that do threaten the health of nearby communities. For example, upstream processes involved in the production and distribution of oil and natural gas emit numerous hazardous air pollutants in addition to methane, including particulate matter (PM), secondary ozone formation, and non-methane volatile organic compounds (VOCs), (Ahmadov et al. 2015; Brantley et al. 2015b; Eisele et al. 2016; Gilman et al. 2013; Helmig et al. 2014; Koss et al. 2017; Roy et al. 2014c) several of which are associated with neurological damage, birth defects, and cancer (Garcia-

Gonzales et al. 2019c; Johnston et al. 2019a). California studies indicate that living in proximity to active oil and gas production wells is associated with increased risk of adverse birth outcomes (Tran et al. 2020b),(Gonzalez et al. 2020b). Air quality sampling during the largest point-source methane release ever recorded in the U.S.–the Aliso Canyon Natural Gas Storage field active blowout in 2015–revealed elevated levels of several hazardous air pollutants including benzene, a carcinogen and reproductive toxicant (Garcia-Gonzales et al. 2019a). Policies aimed at reducing methane emissions also show co-benefits in terms of non-methane VOC and criteria air pollutant emissions. For example, a recent analysis showed that implementing strong federal and state methane policies in the oil and gas sector would result in 1400 fewer deaths and health benefits of $14 billion in 2028 (Buonocore et al. 2021).

Landfills can contaminate local drinking water supplies with hazardous chemicals via leachate, and also release "biogas," an odorous chemical mixture of methane, CO_2, and other VOCs. Residence near landfills has been associated with elevated rates of cancer, low birth weight, and birth defects (Goldberg et al. 1995; Vrijheid 2000).

Research has also documented releases of ammonia, hydrogen sulfide, endotoxins, pathogens, and other airborne contaminants along with methane from CAFOs, and residence near these operations is associated with asthma, decreased lung function, stress, and infection with antibiotic resistant bacteria (Casey et al. 2015a). Several studies report correlations between dairy farm ammonia and greenhouse gas emissions (Miller et al. 2015; Ngwabie et al. 2009; Wu et al. 2012). These releases can further contribute to PM formation and exceedance of National Ambient Air Quality Standards for $PM_{2.5}$ in intense CAFO areas like California's San Joaquin Valley (Eilerman et al. 2016; Miller et al. 2015; Neuman et al. 2003).

Refineries emit hazardous air pollutants, including BTEX compounds (benzene, toluene, ethylbenzene, and xylene), and criteria air pollutants (Mukerjee et al. 2020; Sanchez et al. 2019; Sun et al. 2019); gas power plants may co-emit the same pollutants along with leaked or incompletely combusted methane (Burger et al. 2016; van Kesteren et al. 2013). Such emissions can impact community health, including higher risks of cancer (Yang et al. 2000; Yu et al. 2006) and respiratory problems (Rusconi et al. 2011; Smargiassi Audrey et al. 2009; White et al. 2009).

Methane also contributes to the formation of ground-level ozone, which is linked to premature mortality, impaired respiratory health, and metabolic effects (Jerrett et al. 2009; U.S. Environmental Protection Agency 2020). By one estimate, reducing global methane emissions by 20% would result in approximately 370,000 avoided deaths over twenty years via reductions in global background ozone concentrations (West et al. 2006). Finally, many methane-emitting industries are predominately located in rural communities that also face reduced access to health care, higher rates of poverty, and lower rates of employment compared to urban areas (Kelly-Reif and Wing 2016; Ricketts 2000; Singh and Siahpush 2014). These social stressors may worsen the health effects of pollutant exposures associated with high methane emitters.

In this study, we leverage data from a recent effort to identify high methane emitters in California using airborne remote-sensing (Duren et al. 2019b) and estimates of community demographics refined via novel dasymetric mapping techniques to characterize populations residing near high methane emitters with respect to race, ethnicity, and socioeconomic status

(SES). Our analyses operationalize area-level measures of race/ethnicity and SES to assess inequities in community burdens of high methane emitters and inform strategies to address potential environmental injustices in regulatory enforcement and permitting of these sources of potent greenhouse gases and co-pollutants.

5.3 Materials and Methods

In this cross-sectional environmental justice analysis of high methane emitters in California, we used the block group as our unit of analysis. Prior research indicates this is an appropriate spatial scale to assess racial/ethnic and socioeconomic disparities in environmental exposure (Krieger et al. 2003). All California block groups included in the U.S. Census Bureau's 2016 TIGER/Line Files were eligible for inclusion.

High methane emitter data

We obtained data on high-emitters from the California Air Resources Board (CARB) (Duren et al. 2019b). In brief, CARB provided data from the California Methane Survey conducted by NASA's Jet Propulsion Laboratory, which used Next Generation Airborne Visible/Infrared Imaging Spectrometer (AVIRIS-NG) flights conducted between 2016–2018 to provide systematic information on methane emission point sources. The AVIRIS-NG flights identified 564 distinct strong methane point sources and their average hourly emission rates (kg/hour). The investigators assigned infrastructure elements within energy, agriculture, and waste sectors. From these descriptors, we created seven high-emitters categories: landfill/compost, power plant, refinery, wastewater treatment, oil/gas distribution (i.e., compressors, storage facilities, distribution lines, processing plants, liquid natural gas stations, and gathering lines), oil/gas production (stacks, drill rigs, tanks, lagoons, pump-jacks, plugged wells, and unknown infrastructure), and dairy/manure. We excluded high-emitters located >2km from the boundary of a populated area (n = 81 (14%), **Supplemental Table 1**).

Sociodemographic data

For analyses, we used 2012–2016 American Community Survey data(Manson et al. 2017) to compute block group characteristics: population density (individuals per km^2), percent Hispanic and percent non-Hispanic Native American, Asian, Black, and White, percent rural dwellers, percent linguistically isolated households (i.e., no one in the household older than 14 speaks English "very well"), as well as five measures of SES: percent living below the federal poverty threshold, percent with less than a high school education, percent unemployed, percent renters (vs. home owners), percent Supplemental Nutrition Assistance Program recipients, median household income. Urban block groups consisted of 100% urban population, semi-rural contained >0 to 99% urban population, and rural 0%. A block group-level measure of voter turnout was created using precinct-level elections data from the Statewide Database, California's redistricting database (Statewide Database | Election Data), following Maizlish 2016 (Maizlish 2016). This measure is the average percent of registered voters who voted in the 2012 and 2016 general elections.

High methane emitter exposure measures

To characterize populations living close to high-emitters in California, we constructed a high-resolution spatial layer representing populated areas at sub-block granularity using novel dasymetric mapping methods. Dasymetric mapping refers to the process of disaggregating spatial data – in this case census block boundaries – to finer spatial units of analysis using ancillary data. It has been used in prior environmental justice analyses (Clough and Bell 2016b) and helps to accurately identify residences in rural settings where census blocks (the smallest census geographic unit) can be large (i.e., > 50 km^2) and sparsely populated. Two ancillary data sources were used along with census block population estimates to construct this layer: 1) a statewide database of tax parcel boundaries (smaller than census blocks) from DMP LightBox(Nationwide Parcel Data & Property Level Geocodes | SmartParcels®); and 2) a layer of building footprints for nearly 11 million buildings in California, part of a nationwide layer developed by Microsoft using satellite imagery and machine learning classification techniques (Microsoft/USBuildingFootprints 2019).

Creation of the final populated areas layer using these data followed a tiered process. First, for each census block, we identified all residential parcels within it based on land use descriptions provided in the statewide parcel dataset for each individual parcel (**Supplemental Table 5.2**, **Supplemental Figure 5.1**). If residential parcels were identified in a given block, its population was assumed to be located within these residential areas alone. This parcel-based apportionment accounted for 91.8% of California's population.

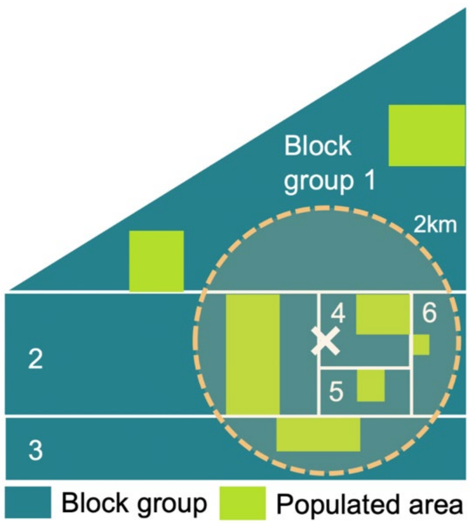
Second, for those blocks containing no residential parcels, but which had a non-zero population count according to the 2010 Census, we allocated population evenly across all building footprint areas identified within them. This was common for sparsely populated blocks in wilderness areas or zones of low-density agriculture, with parcels classified as 'open space' or 'agricultural' in the statewide parcel database, but which still contain residences. Apportioning population to all building footprint areas in these blocks has the advantage of masking out all open land from being considered as populated area but has the disadvantage of misallocating some population to non-residential buildings (e.g., barns, warehouses, processing facilities). This building footprint-based apportionment accounted for 7.9% of California's population.

Finally, a small number of census blocks contained neither residential parcels nor building footprints, but still had a non-zero population count. These blocks were predominantly in very low-density wilderness areas with parcels generally classified as forests/open space and where tree canopies occluded detection of building rooftops via satellite imagery. We assumed that these blocks' populations were evenly distributed across the entire block area. This 'default' method of population apportionment was applied to 0.3% of the state's population. The final populated areas layer was created by merging the results of these three-tiered population apportionment steps into one statewide map.

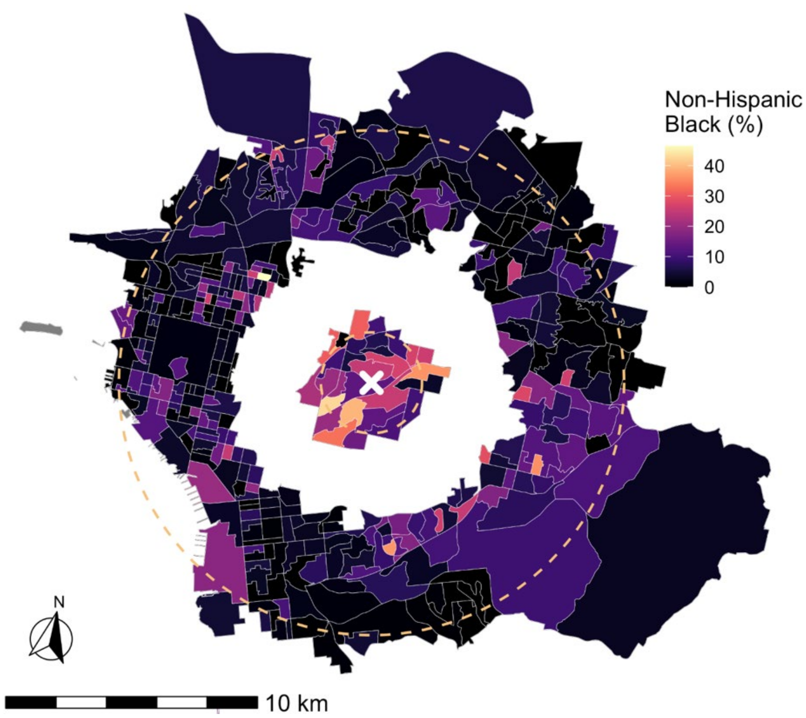
We used the distance between high methane emitters and the dasymetrically mapped populated areas to define exposed and unexposed block groups (**Figure 5.1A**). First, we identified populated areas with boundaries within 2km of a high methane emitter (exposed). Next, we identified all populated areas with boundaries located within 5-10km of a high methane emitter that were also located farther than 5km from all high-emitters (i.e., truly unexposed). Finally, we identified block groups containing the exposed (within 2km of a high methane emitter) and unexposed (5-10km from a high methane emitter) populated areas (**Figure 5.1B**). We opted to define unexposed block groups as those located 5-10km from a high methane emitter in an effort to compare communities similar to the exposed block groups in terms of geographic location, rurality, and other factors, but that differed in high methane emitter exposure status. After removing 24 (0.2%) block groups that were missing sociodemographic data, our study population consisted of 951 exposed and 8,722 unexposed block groups.

Figure 5.1. Example of exposure assignment of block groups. Panel A displays a schematic of block groups (turquoise) and populated areas (light green). Block groups 2-6 are exposed to a high methane emitter (white "X"), but block group 1 is not because its populated areas are located >2km from the high methane emitter . Panel B shows the location of a landfill high methane emitter in San Diego County, California, exposed block groups and the percent of non-Hispanic Black residents within 2km and unexposed block groups 5-10km away. Block groups located 2-4.9km from high-emitters were not included in analysis because we considered them intermediately exposed. The western side of the map crosses over water and thus does not contain block groups. The inner dashed orange line represents the 2km radius around the high methane emitter and the outer dashed orange line the 10km radius around the high methane emitter .

A.



B.



We used two additional metrics to characterize intensity of exposure to high-emitters among block groups located within 2km. We generated a binary multi-category high methane emitter variable that took the value 1 if a block group population area was located within 2km of 2 or more high methane emitter categories (e.g., dairy and oil/gas production) and 0 if a block group population area was located within 2km of a single category of a high methane emitter (e.g., oil/gas distribution only, see **Supplemental Figure 5.2**). We further characterized exposed block groups by the sum of methane emitted from all sources within 2km: $CH_4Exposure_j = \sum_{i=1}^{n} E_i$, where i is a high methane emitter located within 2km of block group j's populated area's boundary and E is the emission rate at high methane emitter i in kg/hour.

Statistical analysis

We conducted descriptive analyses by exposure category. Then we used generalized additive mixed models with a logit-link to assess the association between block group-level sociodemographic variables and odds of exposure to a high methane emitter or, among exposed block groups (those within 2km of a high methane emitter), odds of higher intensity exposure to multiple categories of high-emitters. Mixed models included a random intercept for county. We allowed for deviations from linearity using penalized splines but included a linear term if the generalized cross-validation criterion indicated a linear association was a better fit. We used likelihood ratio testing to select the degrees of freedom for splines. All analyses were conducted using R Statistical Software (Vienna, Austria).

We first ran univariate models, adjusting for population density, for the 14 sociodemographic variables of interest and the three outcomes: 2km vs. 5-10km from a high methane emitter, multiple versus 1 category of high methane emitter exposure, and high versus low CH_4 emissions. We then selected a pared group of variables to include in our fully adjusted models. These variables were selected based on *a priori* hypotheses, e.g., poverty would be associated with high methane emitter exposure, Spearman correlations between the variables, e.g., did not include variables correlated at >0.75 (**Supplemental Figure 5.3**), and associations observed in the univariate models. The adjusted models included: population density, percent individuals of non-Hispanic Asian, Black, and Native American race/ethnicity, and percent individuals of Hispanic race/ethnicity, percent individuals living below the federal poverty threshold, percent voter turnout, percent renters, percent limited English speaking households, and percent uninsured individuals. We used semivariograms to assess residual spatial autocorrelation in our model results(Bivand et al. 2013) and did not observe any (**Supplemental Figure 5.4**).

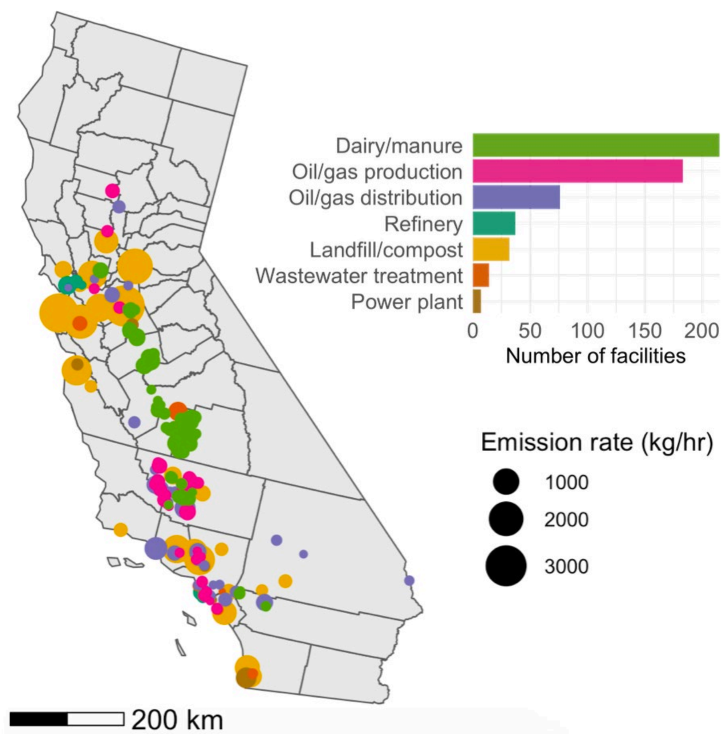
In secondary analyses, we separately assessed the odds of being located within 2km vs. 5-10km from two specific types of high-emitters: (1) oil and gas production; and (2) dairy/manure sites. These two sub-categories of high methane emitter have been associated with environmental justice concerns and adverse health outcomes in prior studies .(Casey et al. 2015b; Donham et al. 2007a; Johnston et al. 2019b; Kroepsch et al. 2019a; Mirabelli et al. 2006a; Wing et al. 2000a)

5.4 Results

AVIRIS-NG flights conducted between 2016–2018 identified 564 high methane emitter in California, 483 (86%) of which we included in analyses as they were located within 2km of a populated area of a block group. **Figure 5.2** shows the spatial distribution of California high-emitters and their relative emission rates. Dairy/manure facilities (N = 213) and oil/gas production sites (N = 127) made up the majority (70%) of the high-emitters. Landfill/compost facilities had the highest emission rates (median [25th, 75th percentile] = 468 kg/hr [254, 1195]) and refineries the lowest (median [25th, 75th percentile] = 20 kg/hr [8, 49], **Supplemental Figure 5.5**). One hundred percent of dairies, 84% of oil and gas production and distribution facilities, and 83% of landfills were in rural or semi-rural block groups while 71% of power plants, 92% of refineries, and 71% of wastewater treatment plants were located in urban block groups.

We identified 951 block groups with populated areas located within 2km of a high methane emitter. Of these, 131 (13.8%) were located within 2km of more than one category of high methane emitter (e.g., a dairy and an oil and gas well). The total hourly methane emissions at high-emitters located within 2km of block groups ranged from 2.8 to 3009 kg/hr (median [25th, 75th percentile] = 93 [40, 185]). The 8,722 block groups located 5-10km from high-emitters constituted our unexposed group.

Figure 5.2. Location, type, and emission rate of high methane emitters (N = 483) in California.



In general, exposed and unexposed block groups had similar sociodemographic characteristics (**Supplemental Table 5.3**). High methane emitter exposed block groups had lower median population density than unexposed block groups (3100 individuals/km^2 versus 4280 individuals/km^2). We observed minimal differences in exposed versus unexposed block groups by high methane emitter category (**Supplemental Figure 5.6**). Larger differences were apparent when comparing number of categories of high methane emitter exposure among exposed block groups, though errors bars were still large (**Supplemental Figure 5.7**). Exposed block groups exposed with 2-4 versus 1 category of high methane emitter, on average, had a higher percentage of Hispanic (50% versus 38%) and a lower percentage of non-Hispanic White individuals (26% versus 39%), a higher percentage of individuals with less than a high school education (26%

versus 19%), and lower voter turnout (63% versus 69%). Similar patterns emerged across categories of total CH_4 emissions exposure within 2km (**Supplemental Figure 5.8**). For example, block groups exposed to high (> tertile 3, 185 kg/hr) versus low (< tertile 1, 40 kg/hr) contained a higher percentage of Hispanic individuals (46% versus 36%), individuals living in poverty (17% versus 13%), linguistically isolated individuals (12% versus 8%), and individuals with less than a high school education (23% versus 16%). We observed strong correlations between several of the sociodemographic variables; e.g., the Spearman ρ between educational attainment and Hispanic race/ethnicity was 0.8, poverty and SNAP use was 0.7, and median household income and poverty was -0.7 (**Supplemental Figure 5.3**).

In unadjusted analyses, we observed multiple non-linear relationships between sociodemographic variables and odds of being located within 2km versus 5-10km from a high methane emitter (**Supplemental Figure 5.9**). For example, as percent non-Hispanic Asian individuals increased, odds of exposure increased, until about 25% non-Hispanic Asians and then there was a steep decline in odds of exposure. The relationship between percent renters and exposure was an inverted U-shape, with the highest odds of being exposed at about 50% renters. Odds of exposure to high-emitters increased linearly with increasing percentage non-Hispanic Black individuals and Native American individuals. We noted somewhat reduced odds of exposure to a high methane emitter with measures of lower SES, except for percent with < high school education. The lowest versus highest population density block groups had three times the odds of being exposed.

In unadjusted analyses considering odds of higher intensity exposure to high-emitters among block groups located within 2km of a high methane emitter, increased odds of exposure to multiple categories of high methane emitter and odds of high exposure to methane emissions was associated with increasing percent Hispanic individuals, uninsured individuals and individuals without a high school diploma (**Supplemental Figures 5.10-5.11**). Increasing percent individuals living in poverty and linguistically isolated households were additionally associated with increased odds of exposure to two or more categories of high methane emitter. While income appeared inversely associated with odds of exposure to two or more categories of high methane emitter, it was positively associated with odds of exposure to high methane emissions. Finally, an increasing percent of non-Hispanic Asian individuals was linearly associated with increased odds of high methane emissions.

When we included 10 sociodemographic variables in a single model, race/ethnicity remained associated with increased odds of being within 2km of a high methane emitter, but SES did not (**Figures 5.3-5.5**). For example, a 10% increase in percent non-Hispanic Black individuals and a 1% increase in non-Hispanic Native American individuals were each associated with a 10% increase in odds (95% CI: 1.04, 1.17 and 1.04, 1.15, respectively) of a block group being located within 2km of a high methane emitter. The associations for non-Hispanic Asian and Hispanic individuals were non-linear. For Hispanics, the relationship was relatively flat until about 50% of the population consisted of Hispanics and then the odds of exposure to a high methane emitter increased (**Figure 5.3**). Once a block group contained 25% non-Hispanic Asians, odds of exposure to a high methane emitter began to decline. Percentage voter turnout demonstrated a unique association with the odds of exposure to a high methane emitter peaking when around 75% of the block group voted and then rapidly declining as that proportion of voters increased.

Figure 5.3. Association between sociodemographic variables and location within 2km versus 5-10km from a high methane emitter. Includes n = 951 exposed and n = 8722 unexposed block groups. Black lines are odds ratios and grey areas represent the 95% confidence intervals. Results from a generalized additive mixed model with a logit link and a random intercept for county adjusted for block group-level percent individuals of non-Hispanic Native American, Asian, and Black race/ethnicity, and percent individuals of Hispanic race/ethnicity, percent individuals living below the federal poverty threshold, percent renters, percent limited English speaking households, percent voter turnout, percent uninsured individuals, and population density. Rug plot displayed along the x-axis shows the number of observations at each level of the respective sociodemographic variable. Non-linear associations in panels B, D, F, H, and J were all statistically significant at the α=0.05 level.
CI, confidence interval; OR, odds ratio.

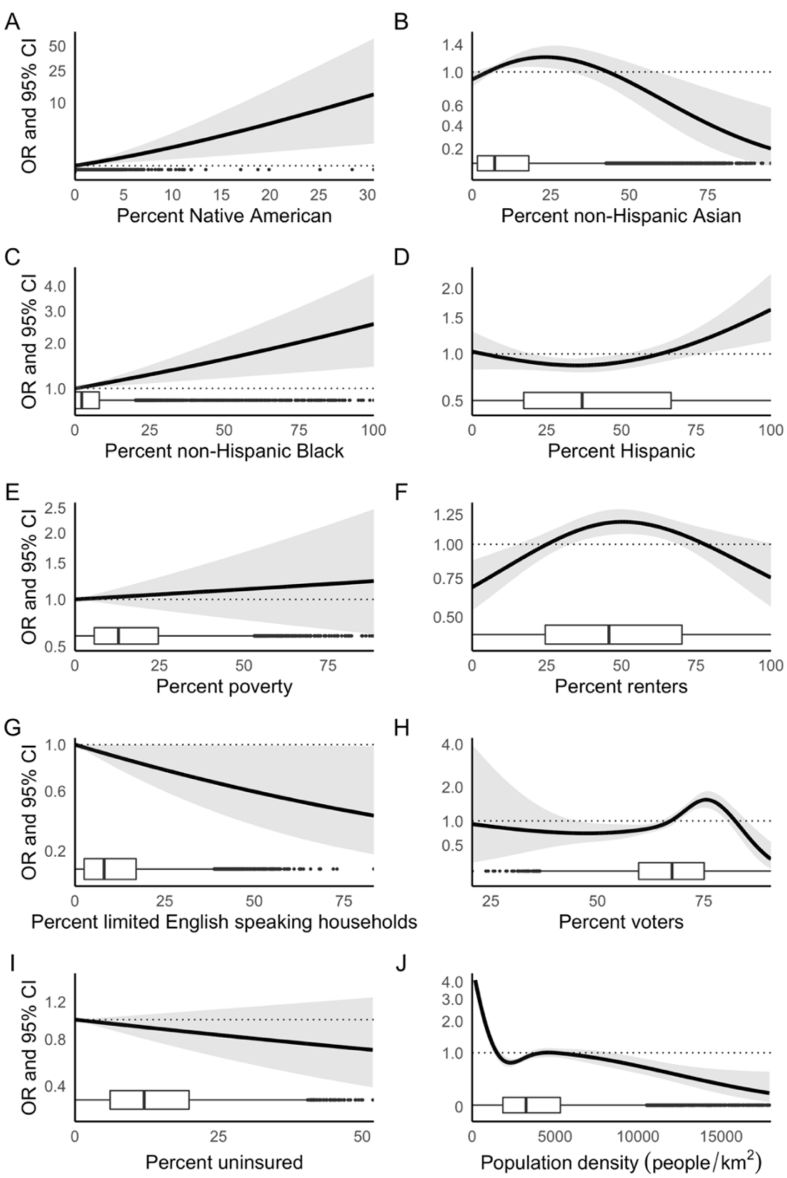


Figure 5.4. Association between sociodemographic variables and location within 2km of 2-4 versus 1 class of high methane emitter, among block groups located within 2km of at least 1 high methane emitter (n = 951). Black lines are odds ratios and grey areas represent the 95% confidence interval. Results from a generalized additive mixed model with a logit link and a random intercept for county adjusted for block group-level for percent individuals of non-Hispanic Native American, Asian, and Black race/ethnicity, and percent individuals of Hispanic race/ethnicity, percent individuals living below the federal poverty threshold, percent renters, percent limited English speaking households, percent voter turnout, percent uninsured individuals, and population density. Rug plot displayed along the x-axis shows the number of observations at each level of the respective sociodemographic variable.
CI, confidence interval; OR, odds ratio. Non-linear associations in panels C and J were statistically significant at the α=0.05 level.

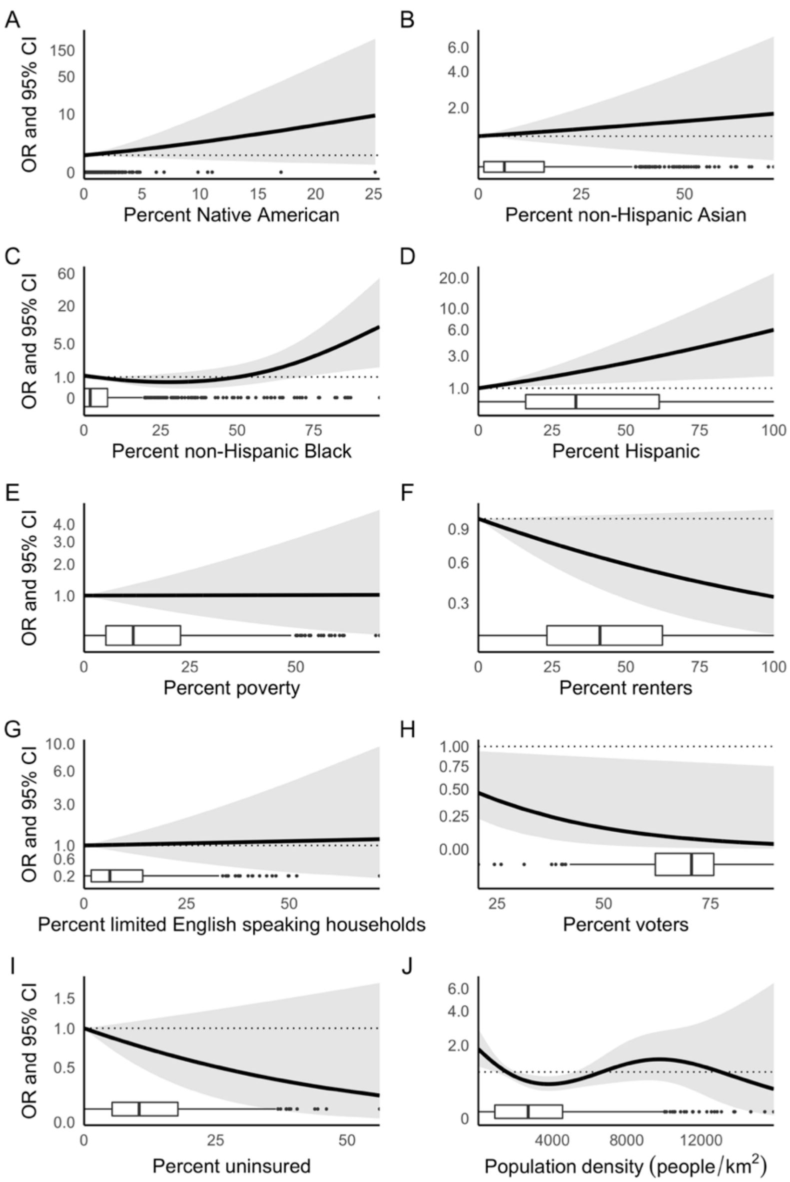
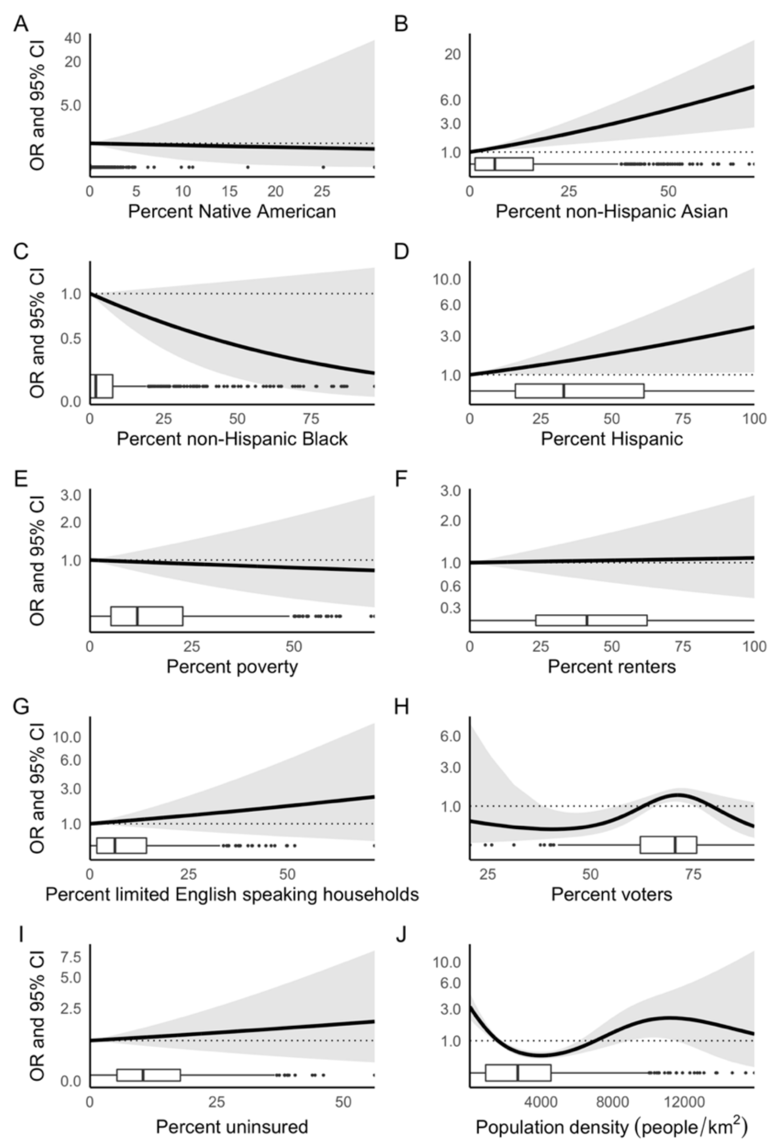


Figure 5.5. Association between sociodemographic variables and exposure to high (>tertile 3 [185 kg/hr]) versus low (tertiles 1-3 [2.8-185 kg/hr]) CH_4 emissions, among block groups located within 2km of at least 1 high methane emitter (n = 951). Black lines are odds ratios and grey areas represent the 95% confidence interval. Results from a generalized additive mixed model with a logit link and a random intercept for county adjusted for block group-level for percent individuals of non-Hispanic Native American, Asian, and Black race/ethnicity, and percent individuals of Hispanic race/ethnicity, percent individuals living below the federal poverty threshold, percent renters, percent limited English speaking households, percent voter turnout, percent uninsured individuals, and population density. Rug plot displayed along the x-axis shows the number of observations at each level of the respective sociodemographic variable. CI, confidence interval; OR, odds ratio. Non-linear associations in panels H and J were statistically significant at the α=0.05 level.



Similar to unadjusted analyses, increasing percent non-Hispanic Black (at 70% Black individuals, OR = 2.33, 95% CI: 0.98, 5.55) and Hispanic individuals (OR = 1.19, 95% CI: 1.04, 1.36 for each 10% increase in Hispanic individuals) were associated with increased odds of being exposed to two or more categories of high methane emitter among block groups located within 2km of a high methane emitter in adjusted analyses (**Figure 5.4**). Increasing percent renters (OR = 0.90, 95% CI: 0.80, 1.01 for each 10% increase) and voter turnout (OR = 0.67, 95% CI: 0.48, 0.95 for each 10% increase) were inversely associated with odds of exposure to two or more categories of high methane emitter. Non-Hispanic Black race/ethnicity was inversely associated with odds of high methane emissions (OR = 0.85, 95% CI: 0.69, 1.03 for each 10% increase in non-Hispanic Black individuals), while increasing percent non-Hispanic Asian (OR = 1.35, 95% CI: 1.14, 1.59) and Hispanic (OR = 1.14, 95% CI: 1.01, 1.29) individuals were associated with increased odds of high methane emissions among block groups within 2km of a high methane emitter (**Figure 5.5**).

When we assessed the odds of being located within 2km of an oil and gas production or a dairy/manure high methane emitter, we observed similar racial/ethnic disparities to those observed for high-emitters overall, with some differences (**Supplemental Figures 5.12-5.13**). For oil and gas production sites, we observed increased odds of exposure with increasing percent Native American, non-Hispanic Black, and non-Hispanic Asian populations. For example, for each 10% increase in non-Hispanic Asian individuals there was a 26% increase in the odds of being located within 2km vs. 5-10km of an oil and gas production high methane emitter (OR = 1.26, 95% CI: 1.06, 1.50). For dairy/manure sites, odds of exposure increased with higher percentages of Native American, Hispanic, and non-Hispanic Black individuals, for whom we observed the strongest relationship (OR = 1.81, 95% CI: 1.07, 3.06) for each 10% increase in non-Hispanic Black individuals).

5.6 Discussion

We examined the location of 483 high methane emitters in relation to community-level demographics based on race/ethnicity, SES, and civic engagement capacity. To our knowledge, this is the first environmental justice analysis to assess relationships between community characteristics and proximity to and intensity of exposure to multiple high methane emitter types, including landfills/composting facilities, power plants, refineries, wastewater treatment plants, oil and gas distribution and production sites, and dairies/manure management sites. Landfills and composting facilities accounted for the highest rates of methane emissions, while dairies and manure management sites as well as oil and gas production facilities made up the largest proportion of high-emitters in our analysis (Duren et al. 2019b). Unadjusted models showed racial/ethnic and SES disparities in the odds of living in close proximity to high methane emitters and intensity of exposure based on multiple industry categories and total methane emissions. In adjusted models, the associations with race/ethnicity persisted, while those for community-level SES (poverty rate, percent uninsured, and percent limited English-speaking households), were attenuated. Further, sub-analyses restricted to dairies/manure management facilities and oil and gas production revealed similar racial disparities as the main analysis. Our sub- and overall analyses also showed many non-linear relationships. Interestingly, once voter turnout, an indicator of community civic engagement, reached 75% the odds of being exposed to a high methane emitter declined. This finding supports the idea that marginalized communities may be vulnerable to siting of environmental hazards due to lack of political power and limited resources to engage in regulatory decision-making or challenge facility permits (Morello-Frosch 2002; Wing et al. 2000b). In addition, 84% of the high methane emitters included in our study were located in semi-rural or rural block groups, highlighting what some researchers argue is an under-studied form of rural environmental injustice in which urban areas drive the intensity of food and energy production in rural areas, and often return their wastes to these same rural communities (Kelly-Reif and Wing 2016). Our results indicate that future methane emission reduction policies to slow climate change, can also address exposure disparities to health-harming co-pollutants. This could be done by prioritizing and incentivizing deeper methane emissions reductions in environmental justice communities.

Prior studies have examined equity patterns of specific sources of methane emissions included in our analysis. For example, US studies of solid and hazardous waste landfills indicate their disproportionate siting in communities of color (Martuzzi et al. 2010; Saha and Mohai 2005). This body of work includes environmental justice assessments of CAFOs showing that weak regulations have led to the disproportionate location of swine CAFOs in communities of color and poor communities (Donham et al. 2007b; Ladd and Edward 2002; Wilson et al. 2012; Wing et al. 2000b) and near schools with predominantly low-income and nonwhite students (Mirabelli et al. 2006b). None of these studies, however, has examined CAFO sites, such as dairies, in California. Our results showed that odds of exposure (within 2km) to this category of high methane emitters tended to increase with increasing percent Native American, non-Hispanic Black, and Hispanic individuals. In contrast to studies of all CAFOs, we did not observe increased odds of exposure among lower SES communities in adjusted models.

Similarly, environmental inequities associated with California's oil and gas industry, in particular production sites, emerged in large part due to historical redlining beginning in the late 1930s through the late 1960s, which restricted many African Americans and Latino immigrant home-buyers to the petro-industrial neighborhoods of South Los Angeles (Cumming 2018; Viehe 1981). This legacy shapes present day race- and class-based inequities in the "petro-riskscapes" of Los Angeles and rural communities in San Joaquin and Kern Counties–epicenters of California's oil and gas production (Srebotnjak, Tanja and Rotkin-Ellman 2014). Our data support this theory. We observed increased odds of being located within 2 versus 5-10km from an oil and gas production high methane emitter with increasing percent Native American, non-Hispanic Asian, and non-Hispanic Black individuals. In addition, the proliferation of unconventional oil and gas extraction technologies, such as hydraulic fracturing, raises new concerns regarding methane emissions (Howarth et al. 2011) and community health effects (Elliott et al. 2017; Garcia-Gonzales et al. 2019c). These sites tend to be located in low income rural communities, such as the Marcellus Shale in Pennsylvania or the Eagle Ford Shale in Texas, and the few environmental justice studies conducted on unconventional drilling indicate that this development is often, though not always, disproportionately located in communities with lower home values and minority communities (Clough and Bell 2016b; Johnston et al. 2016b, 2020b; Kroepsch et al. 2019b; Malin and DeMaster 2016; Ogneva-Himmelberger and Huang 2015b). Strong federal and state methane emission regulations will also reduce non-methane VOC and criteria air pollutant emissions, and such policies have the potential to prevent 1400 deaths and 50,000 asthma exacerbations in 2028 (Buonocore et al. 2021).

This study has several strengths. First, this is the first environmental justice analysis of high methane emitters using several exposure metrics, including proximity to multiple sites, as well as airborne, remotely sensed estimates of cumulative methane emissions from diverse sources. Second, we used a high-resolution dataset of populated areas developed via dasymetric mapping to spatially characterize the location of populations within exposed and unexposed block groups. Third, we examined several demographic variables to assess patterns of inequity, including voter turnout, an indicator of community voice and political power that may be an important driver of environmental justice outcomes. Finally, we used splines and adjusted analyses to allow us to assess nonlinear trends and better isolate which community sociodemographic variables most likely explained observed associations. We found race/ethnicity better predicted exposure than low SES, potentially indicating housing discrimination, segregation, or procedural environmental injustice as drivers (Fernandez-Bou et al. 2021; Mohai and Bryant 1992). We also highlight rurality as an important, yet understudied dimension of environmental injustice in California (Kelly-Reif and Wing 2016).

Limitations include the cross-sectional design, which precludes assessment of temporal changes in block group demographic composition or distributional patterns of cumulative methane emissions; indeed, identification of high methane emitters took place between 2016 and 2018 and emissions trends likely vary over time. In addition, although studies indicate that harmful compounds are often co-emitted with methane (Buonocore et al. 2021; Garcia-Gonzales et al. 2019c; Johnston et al. 2019a), which itself does not directly harm human health, we did not directly measure these co-pollutant emissions, and thus cannot characterize the potential health implications of these sites, which likely vary by high methane emitter category and facility, for host communities. Finally, we treated each facility as a point location even though some facilities, such as dairies, span larger areas. This may have resulted in underestimation of exposed populations.

Future research should reassess temporal fluctuations in methane emissions from high methane emitter sites and the extent to which these emissions correlate with potentially harmful co-pollutants across all facility types. Given that 10% of high-emitters in California were estimated to have contributed roughly 60% of point-source methane emissions, (Duren et al. 2019b) more targeted air quality monitoring, in collaboration with host communities, could provide much-needed data to better understand potential community health threats posed by these sites. While some analysts have cautioned against integrating air quality into climate policy, pointing out that co-pollutants are best regulated under existing laws such as the US Clean Air Act, (Schatzki and Stavins 2009) more holistic regulatory strategies could target critical methane emission reductions to those communities where health co-benefits and health equity impacts are greatest (Boyce and Pastor 2013; Shonkoff et al. 2011). California's Assembly Bill 617 (Garcia 2017) provides an innovative and potentially transformational blueprint for enhanced community participation in air monitoring and development of emissions reduction plans to improve local air quality and ultimately reduce environmental health disparities in disadvantaged communities (Community Air Protection Blueprint | California Air Resources Board). This legislative strategy to localize air quality management from a regional scale to a community scale can also embed environmental justice objectives in efforts to identify and more effectively regulate high methane emitters. Indeed, harmonizing environmental justice and climate sustainability goals to incentivize greenhouse gas reductions in disadvantaged and highly polluted neighborhoods could enhance overall health benefits, particularly if a small number of high methane emitters present the greatest opportunities to improve local and regional air quality. This would require systematic temporal and spatial tracking of methane and co-pollutant emissions to characterize the health and environmental justice implications of high-emitters more fully. Such a strategy would also advance the overarching environmental justice goals articulated in California's landmark climate change laws.

5.7 Supplemental Information Chapter 5

Supplemental Table 5.1. High-emitters excluded and included in analyses.

Category	**No populated area within 2km (excluded)**	**Populated area within 2km (included)**
Landfill/compost	3 (9%)	29 (91%)
Power plant	0	7 (100%)
Refinery	0	37 (100%)
Wastewater treatment	0	14 (100%)
Oil/gas distribution	20 (26%)	56 (74%)
Oil/gas production	56 (31%)	127 (69%)
Dairy/manure	2 (1%)	213 (99%)
Total	81 (14%)	483 (86%)

Supplemental Table 5.2. Residential Parcel Classifications. Residential parcels were defined as any parcel classified by one of the following land-use codes in the statewide tax parcel database. Large parcels for any use low-density use code with areas greater than 1-acre (4,047 m^2) were assumed to contain unpopulated, open space and were excluded. High-density residential parcels that tend to be larger (e.g. an entire apartment complex) were allowed to have areas of up to 50-acres before being excluded and are indicated in the list below by (*).

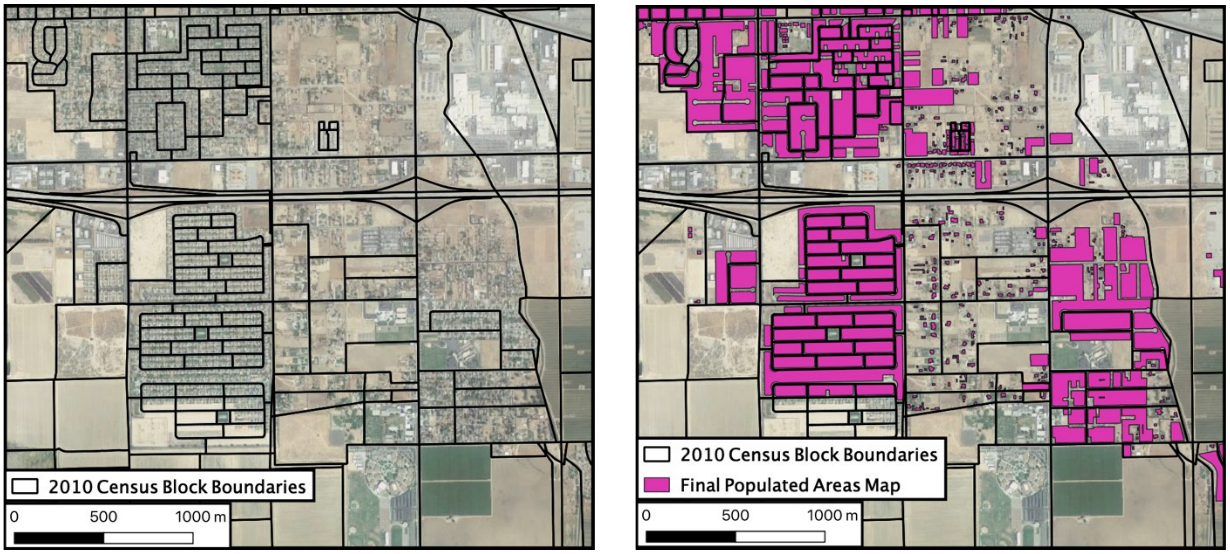
Parcel Land-Use Code
APARTMENT HOUSE (100+ UNITS)*
APARTMENT HOUSE (5+ UNITS)*
APARTMENTS (GENERIC)*
CLUSTER HOME (RESIDENTIAL)
COMM/OFC/RES MIXED USE
CONDOMINIUM (RESIDENTIAL)*
COOPERATIVE (RESIDENTIAL)*
DORMITORY, GROUP QUARTERS (RESIDENTIAL)
DUPLEX (2 UNITS, ANY COMBINATION)
FRATERNITY HOUSE, SORORITY HOUSE
GARDEN APT, COURT APT (5+ UNITS)*
HIGHRISE APARTMENTS*
HOMES (RETIRED; HANDICAP, REST; CONVALESCENT; NURSING)
MANUFACTURED, MODULAR, PRE-FABRICATED HOMES
MISC RESIDENTIAL IMPROVEMENT
MOBILE HOME
MOBILE HOME PARK, TRAILER PARK
MULTI-FAMILY DWELLINGS (GENERIC, ANY COMBINATION 2+)
PLANNED UNIT DEVELOPMENT (PUD) (RESIDENTIAL)
QUADRUPLEX (4 UNITS, ANY COMBINATION)
RESIDENTIAL (GENERAL) (SINGLE)
RESIDENTIAL COMMON AREA (CONDO/PUD/ETC.)
RESIDENTIAL INCOME (GENERAL) (MULTI-FAMILY)
RURAL RESIDENCE (AGRICULTURAL)
SINGLE FAMILY RESIDENTIAL
STORES & APARTMENTS
TIMESHARE (RESIDENTIAL)
TOWNHOUSE (RESIDENTIAL)
TRIPLEX (3 UNITS, ANY COMBINATION)
ZERO LOT LINE (RESIDENTIAL)

Supplemental Table 5.3. Distribution of sociodemographic variables by exposed versus unexposed groups and scaling factors used for regression analyses.

	Exposed block groups (within 2km of a high methane emitter) N = 951	**Unexposed block groups (5-10km from a high methane emitter) N = 8,722**	**Linear ß interpretation**
Category	Median (25^{th}, 75^{th} percentiles)		
Race/ethnicity, %			
Hispanic	33 (16, 61)	37 (17, 67)	Per 10% increase
Non-Hispanic			
Native American[a]	0.4 (1.7)	0.2 (0.9)	Per 1% increase
Asian	6 (1, 16)	7 (2, 18)	Per 10% increase
Black	2 (0, 8)	2 (0, 8)	Per 10% increase
White	35 (11, 61)	28 (8, 56)	N/A
Poverty, %	15 (13, 23)	17 (14, 25)	Per 10% increase
Renters, %	41 (23, 62)	47 (25, 71)	Per 10% increase
Limited English-speaking households, %	6 (2, 14)	8 (3, 17)	Per 1% increase
Voters, %	71 (62, 76)	67 (59, 75)	Per 10% increase
Uninsured, %	10 (5, 18)	12 (6, 20)	Per 1% increase
Median household income, $	$64,700 (45,000, 92,200)	$60,900 (41,700, 86,300)	Per $10,000 increase
Less than a high school diploma, %	14.2 (5.3, 31.4)	16.2 (6.0, 33.5)	Per 10% increase
Unemployed, %	8 (5, 12)	8 (5, 12)	Per 1% increase
SNAP, %	7 (1, 15)	7 (2, 17)	Per 1% increase
Population density, individuals per km^2	3100 (2750, 4540)	4280 (3820, 5510)	Per 500 individuals per km^2 increase

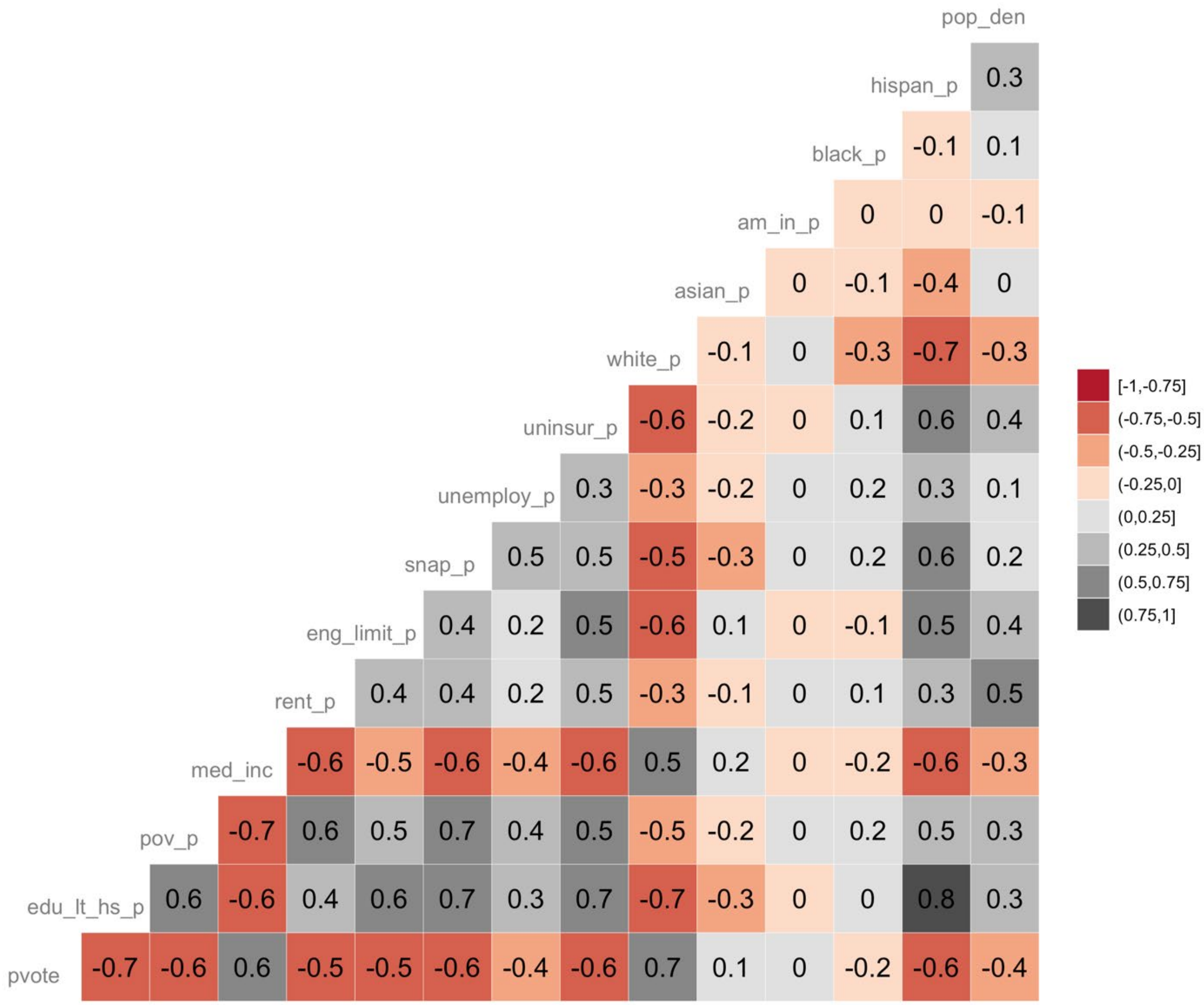
[a] Mean (SD)

Supplemental Figure 5.1. Example of the creation of populated areas layer from parcel, building footprint and block boundary data. Image shown is in eastern Bakersfield, CA.

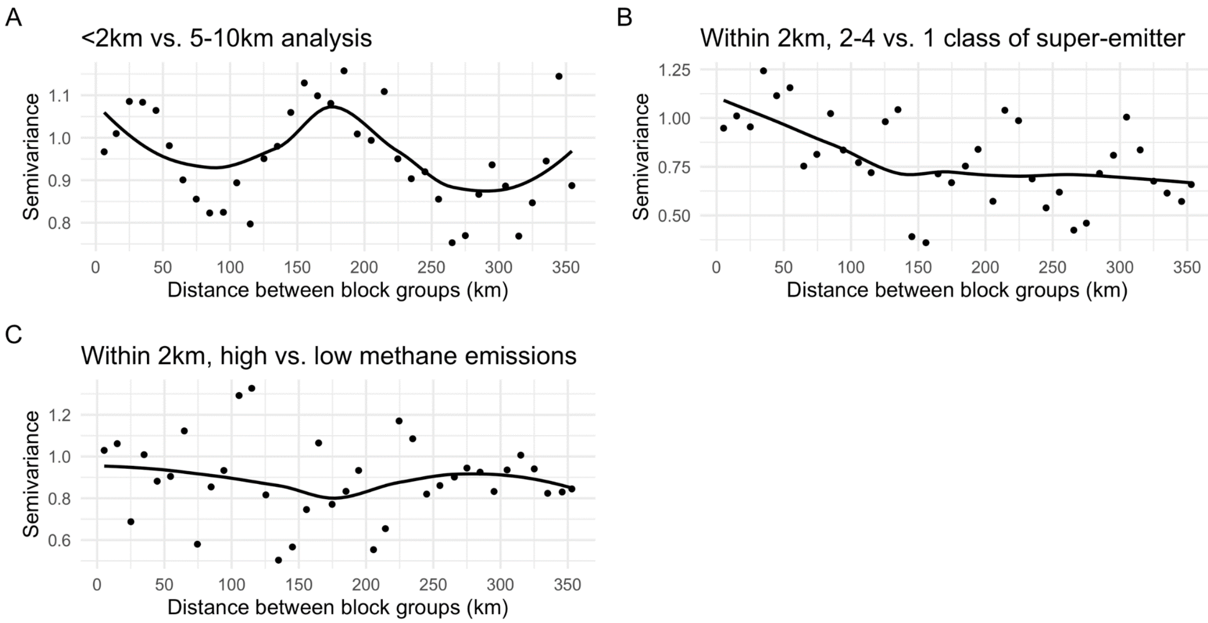


Supplemental Figure 5.2. Example of block groups exposed to multiple classes of high methane emitter. For example, the block group outlined in a dashed white line contains populated areas located within 2km of at least two classes of high methane emitter (wastewater treatment and oil and gas distribution). White hexagons represent wastewater treatment facilities, squares are oil and gas production sites, circles are oil and gas distribution sites, and triangles are landfills. Grey polygons are populated areas within block groups and the larger polygons bounded in black are block groups. Analyses were conducted at the block group-level but only those block groups with a populated area located within 2km of a high methane emitter were included.
O&G, oil and gas.

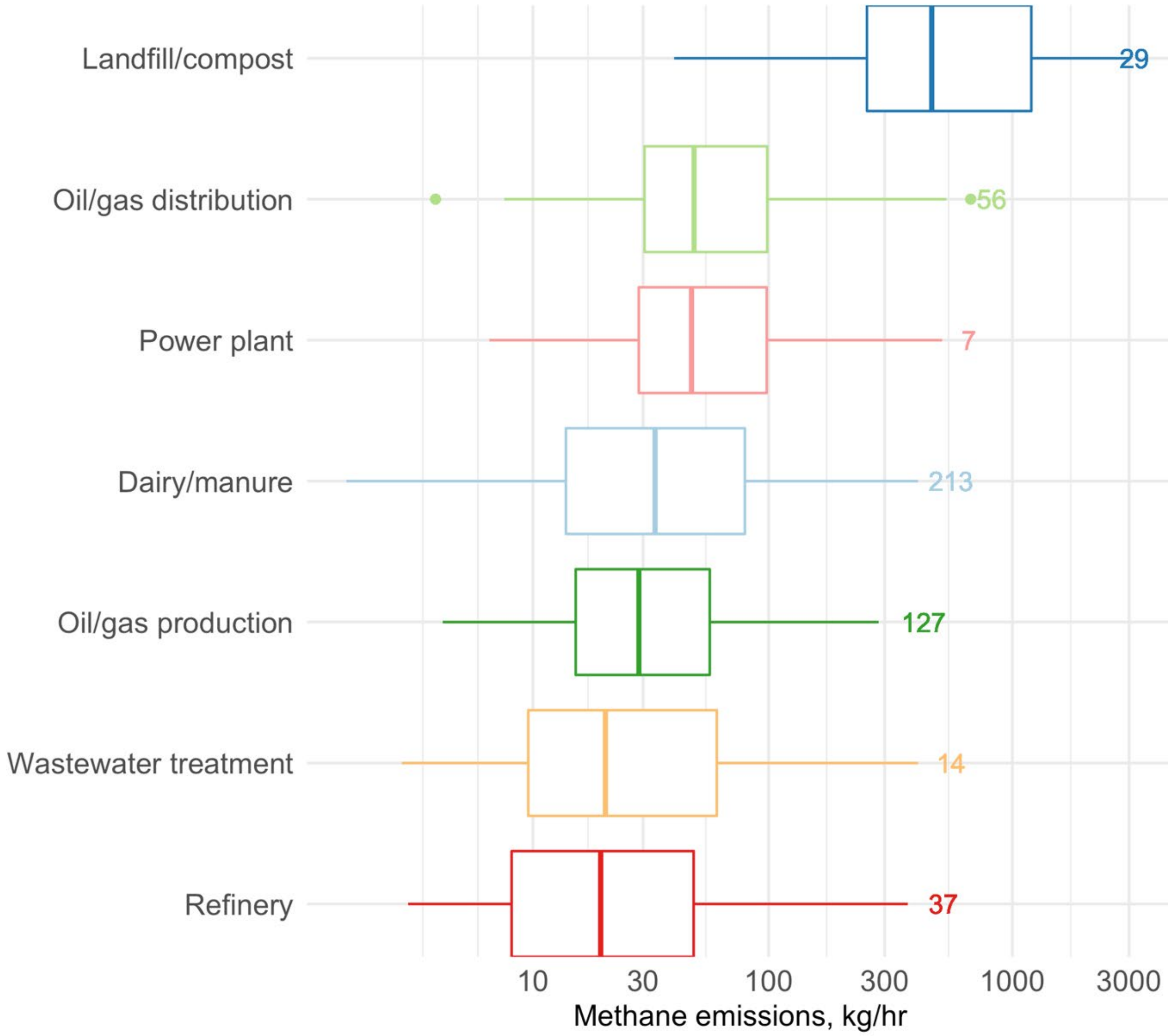
Supplemental Figure 5.3. Spearman correlation matrix for block group-level sociodemographic variables.



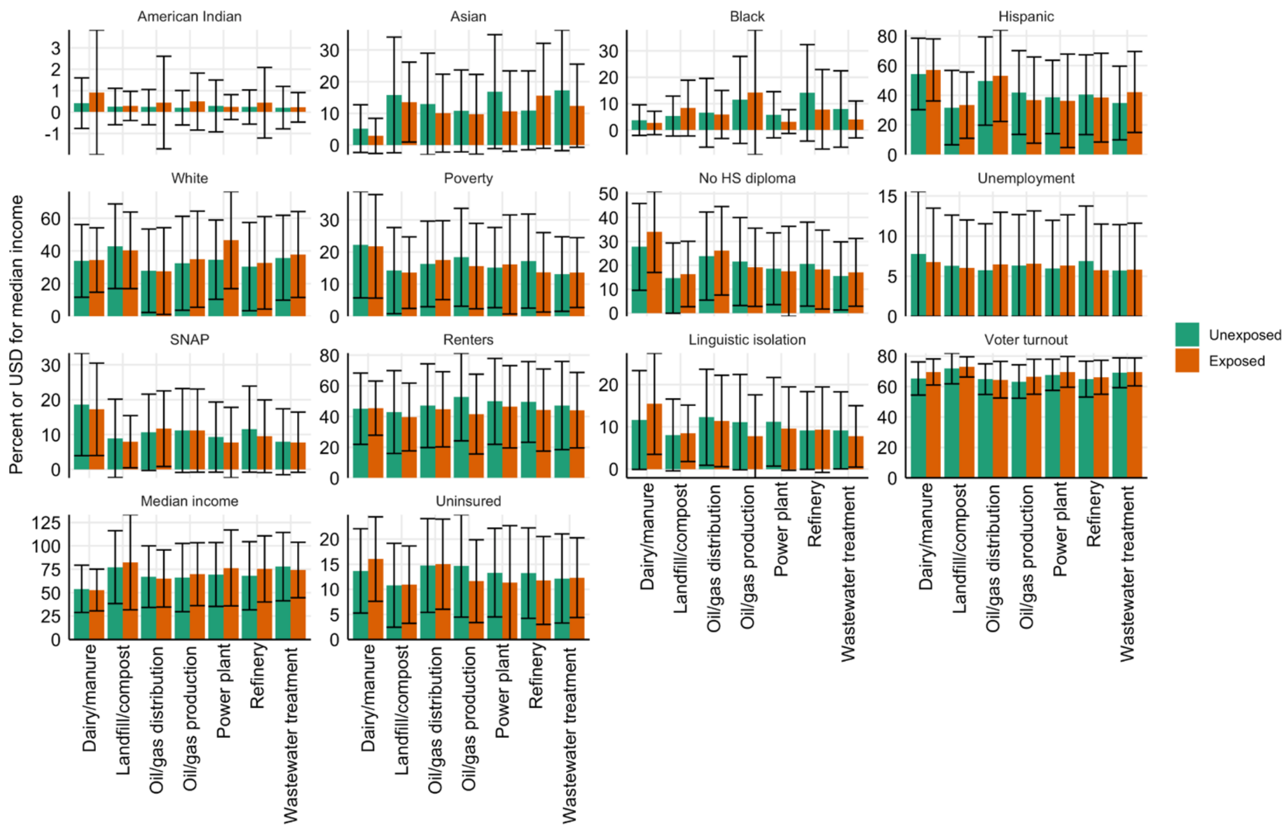
Supplemental Figure 5.4. Semivariograms for the three analyses: A. Main 2km vs. 5-10km; B. 2-4 vs. 1 class of high methane emitter within 2km; C. High (>3rd quartile) vs. low (quartiles 1-3) CH_4 emissions within 2km. The shapes of the semivariograms are consistent with limited residual spatial autocorrelation. Based on residuals from logistic mixed models with a random intercept for county adjusted for block group-level for population density, percent individuals of non-Hispanic Asian, Black, and Native American race/ethnicity, and percent individuals of Hispanic race/ethnicity, percent individuals living below the federal poverty threshold, percent voters, percent renters, percent limited English speaking households, and percent uninsured individuals.



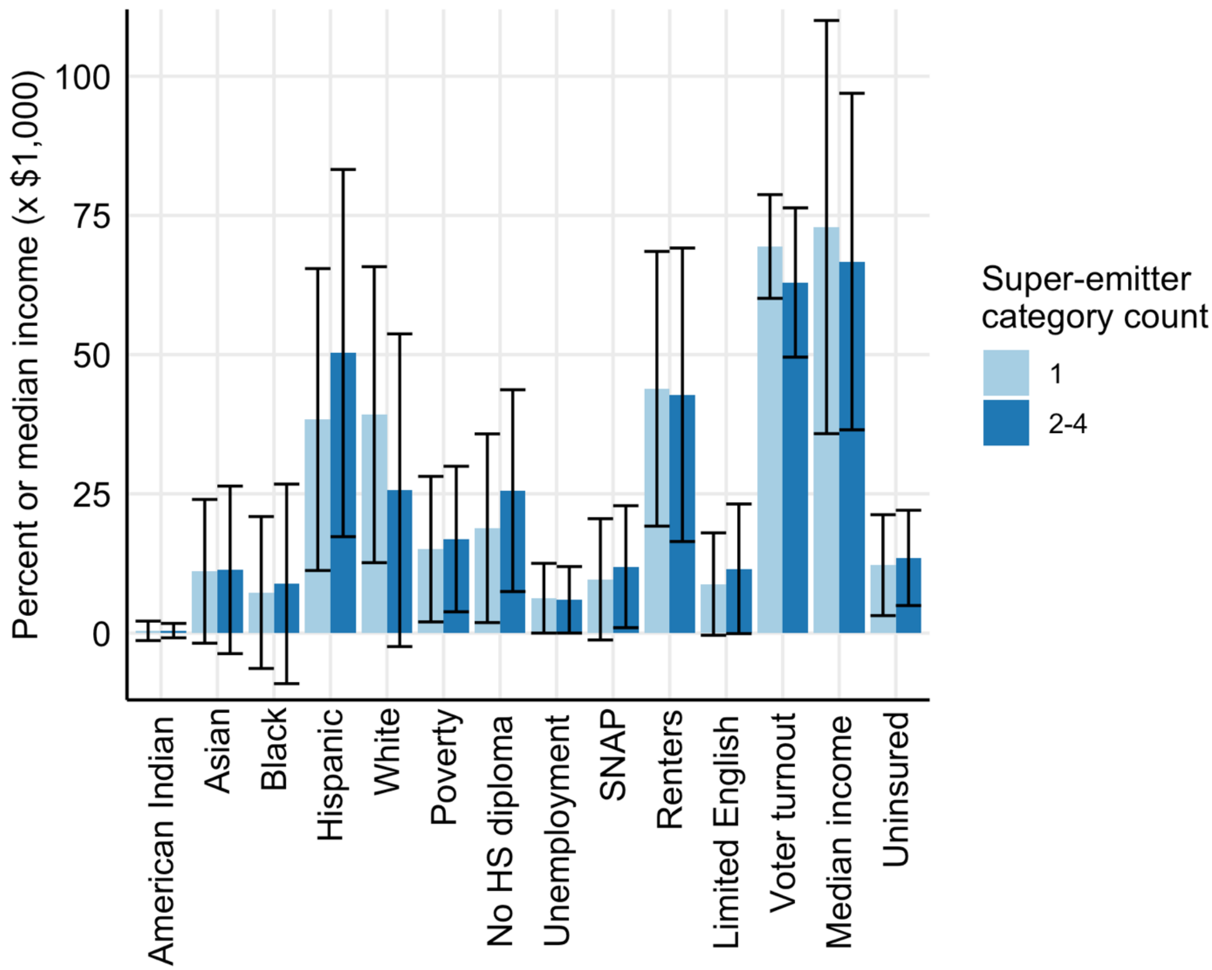
Supplemental Figure 5.5. Distribution of methane emissions (kg/hr) by high methane emitter category. Numbers indicate the count of high-emitters in each category. The x-axis is log-scale.



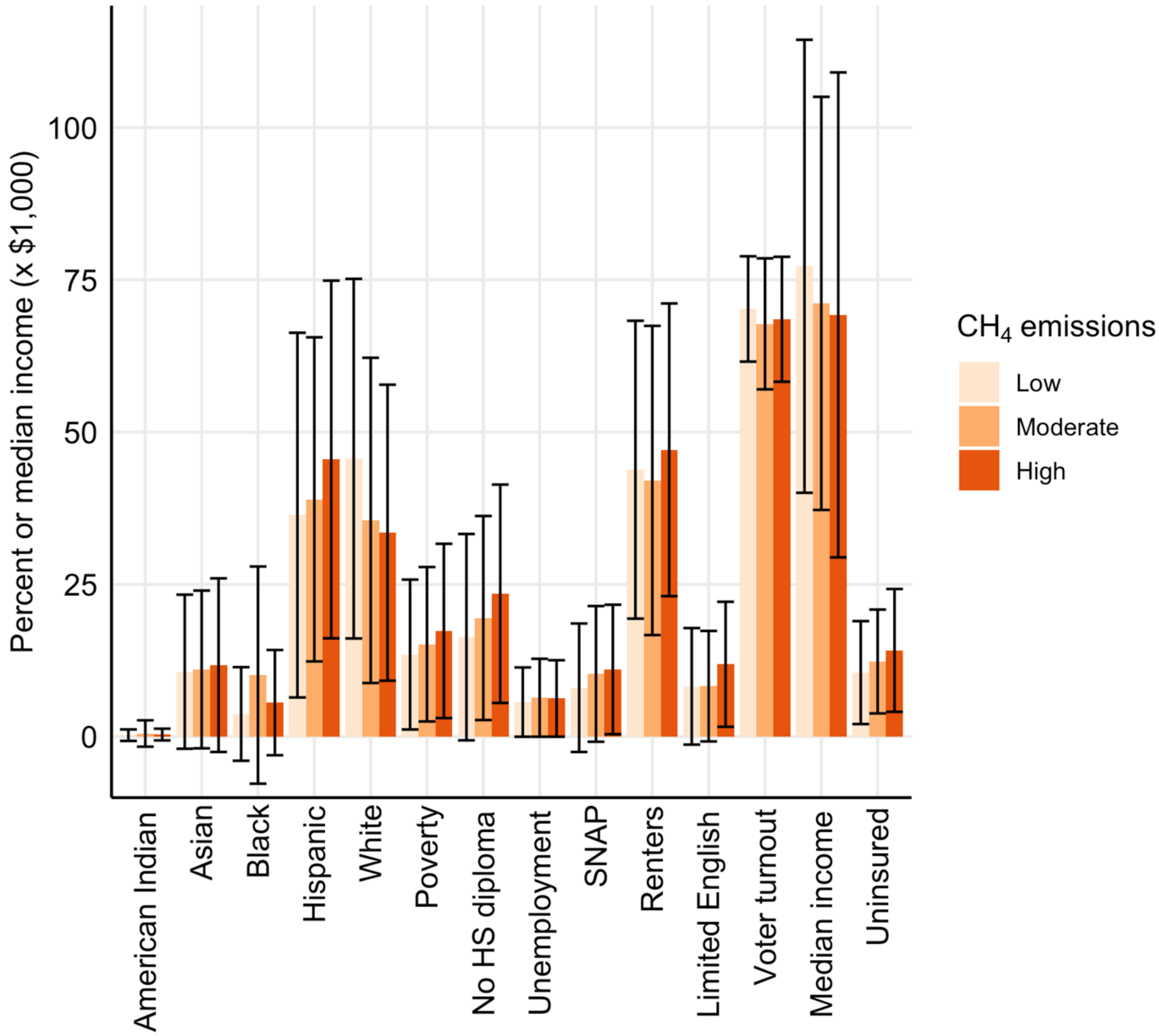
Supplemental Figure 5.6. Average block group-level sociodemographic characteristics by high methane emitter class. Exposed block groups were those with a populated area located within 2km of a high methane emitter and unexposed those located 5-10km from a high methane emitter .
Bars represent 1-SD.



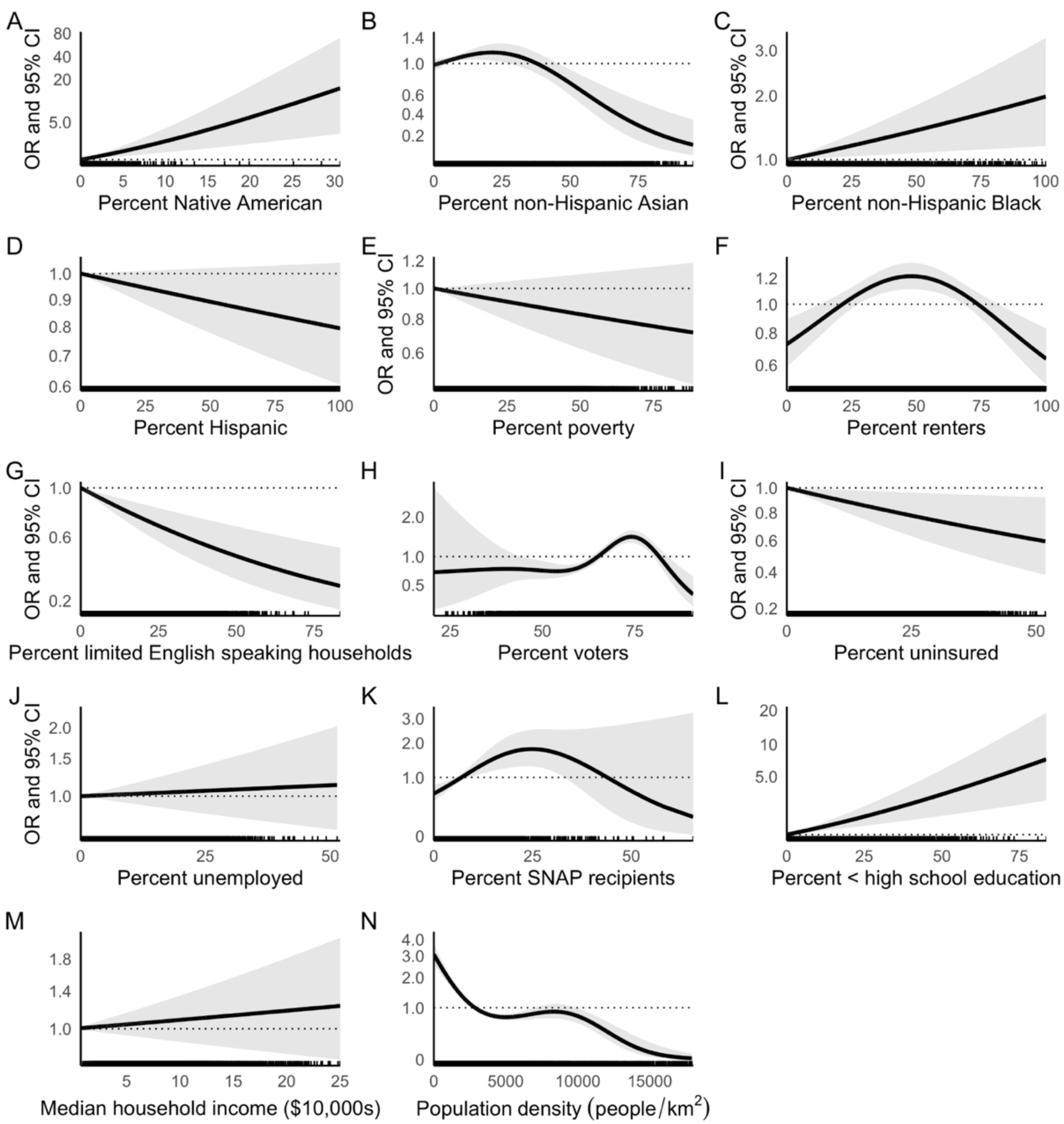
Supplemental Figure 5.7. Block group-level sociodemographic characteristics among block groups located within 2km of a high methane emitter , stratified by the number of categories of high methane emitter located within 2km. For example, block groups located within 2km of a refinery and a dairy would fall in the 2-4 category.
Bars represent 1-SD.



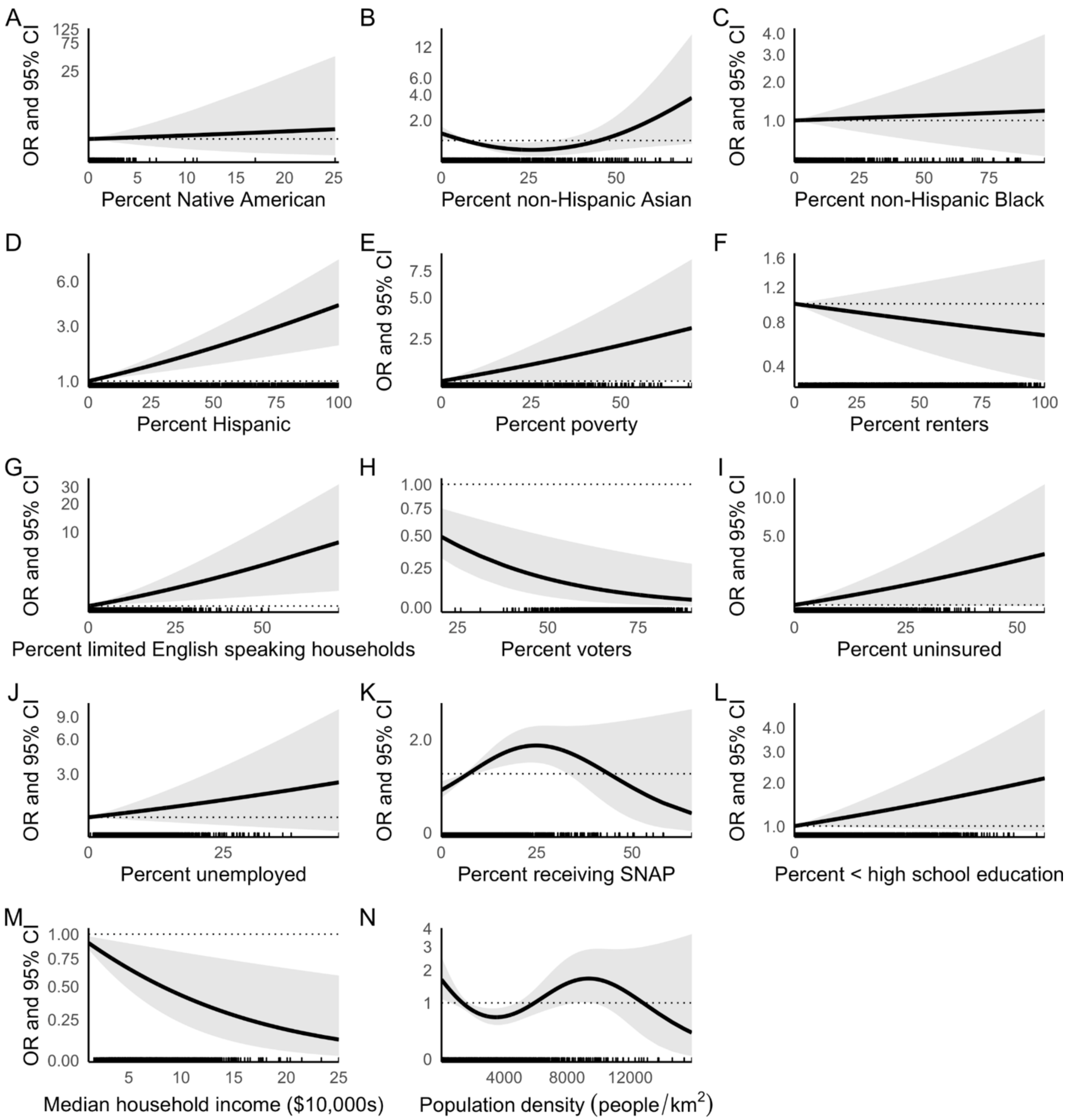
Supplemental Figure 5.8. Block group-level sociodemographic characteristics by the sum of high methane emitter CH_4 emissions (kg/hr) within 2km of the block group. CH_4 emissions were categorized based on their distribution into low (<40 kg/hr, first quartile), moderate (40 to <185/hr, third quartile), and high (>185 kg/hr).



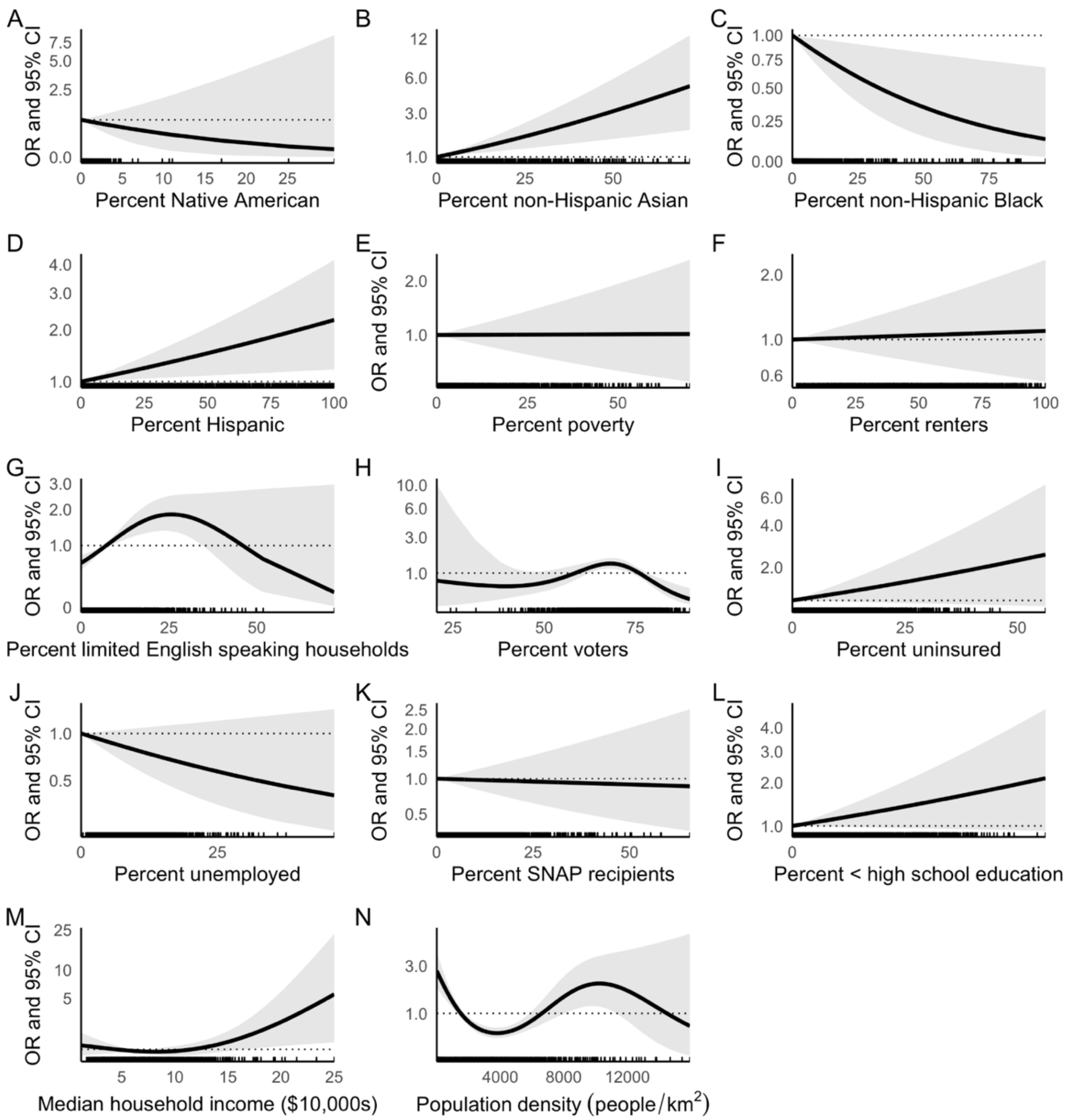
Supplemental Figure 5.9. Unadjusted association between sociodemographic variables and odds of being located within 2km versus 5-10km from a high methane emitter . Includes n = 951 exposed and n = 8722 unexposed block groups. Black lines are odds ratios and grey areas represent the 95% confidence intervals. Results from a generalized additive mixed model with a logit link and a random intercept for county adjusted only for block group-level population density. Rug plot displayed along the x-axis shows the number of observations at each level of the respective sociodemographic variable.
CI, confidence interval; OR, odds ratio. Non-linear associations in panels B, F, H, K, and N were statistically significant at the α=0.05 level.



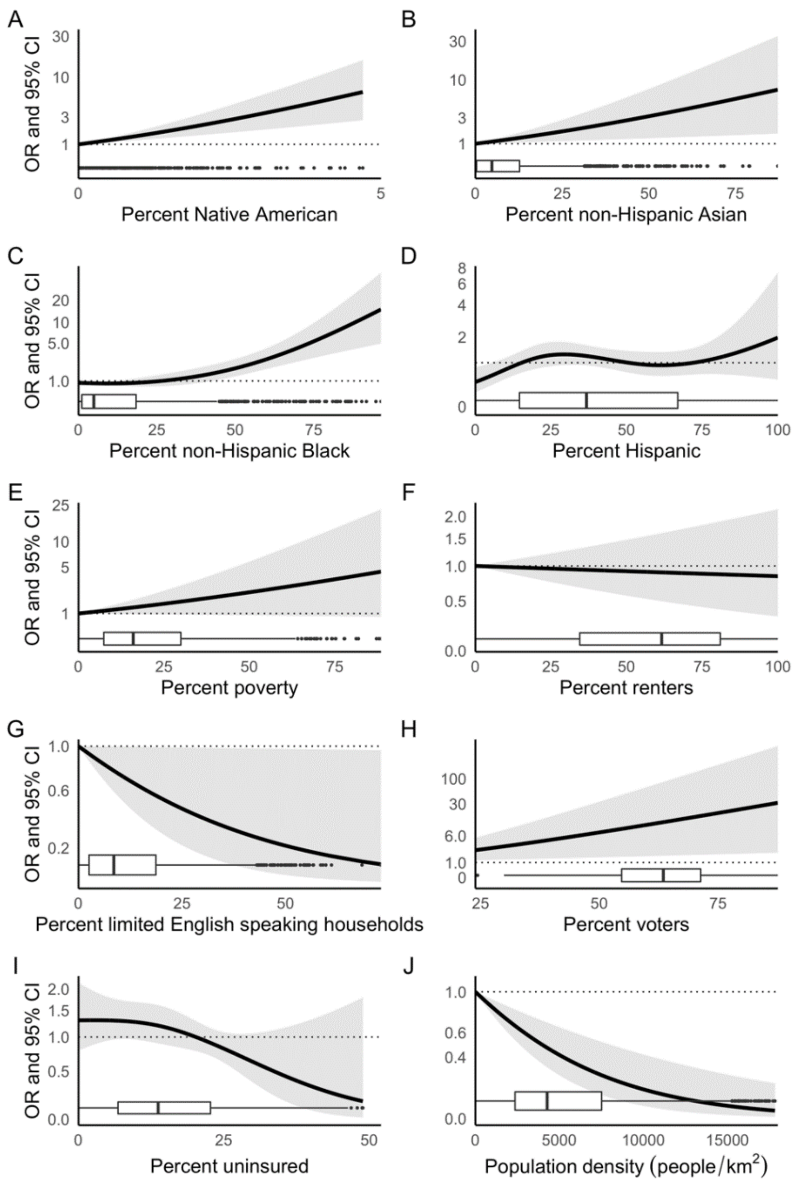
Supplemental Figure 5.10. Unadjusted association between sociodemographic variables and odds of being located within 2km of 2-4 versus 1 category of high methane emitter , among block groups located within 2km of at least 1 high methane emitter (n = 951). Black lines are odds ratios and grey areas represent the 95% confidence interval. Results from a generalized additive mixed model with a logit link and a random intercept for county adjusted only for population density. Rug plot displayed along the x-axis shows the number of observations at each level of the respective sociodemographic variable.
CI, confidence interval; OR, odds ratio. Non-linear associations in panels B, K, and N were statistically significant at the α=0.05 level.



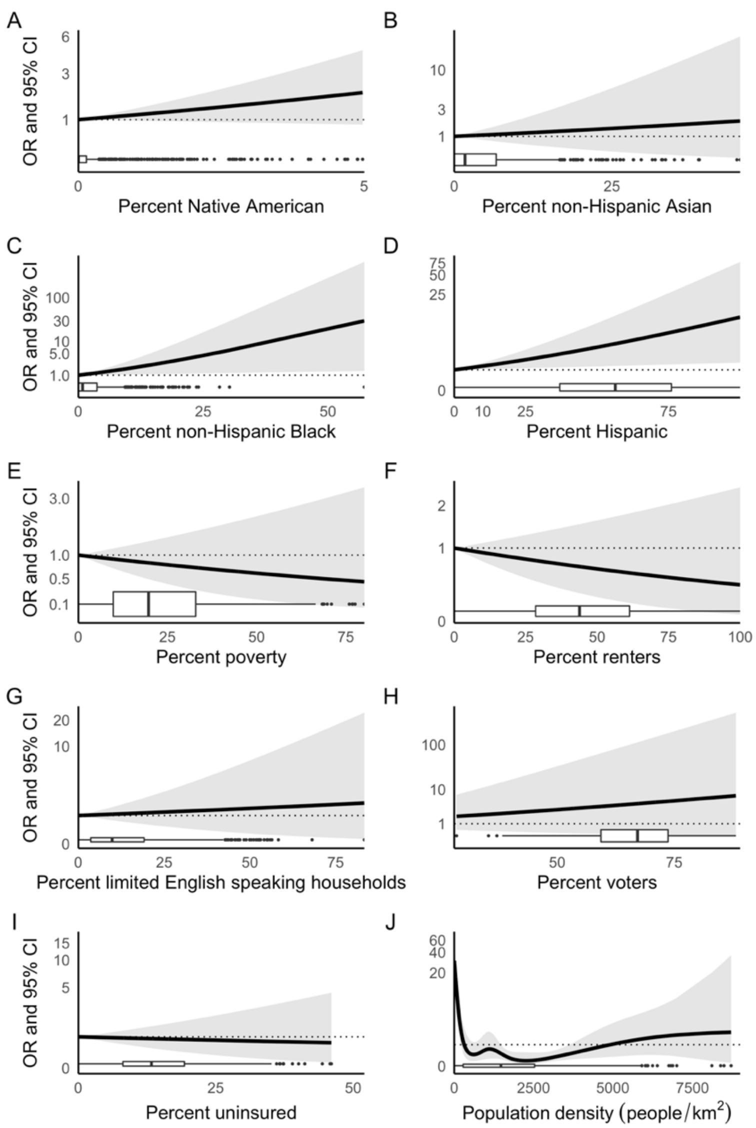
Supplemental Figure 5.11. Unadjusted association between sociodemographic variables and odds of being exposed to high (>quartile 3 [185 kg/hr]) versus low (quartile 1-3 [2.8-185 kg/hr]) CH_4 emissions, among block groups located within 2km of at least 1 high methane emitter (n = 951). Black lines are odds ratios and grey areas represent the 95% confidence interval. Results from a generalized additive mixed model with a logit link and a random intercept for county adjusted only for population density. Rug plot displayed along the x-axis shows the number of observations at each level of the respective sociodemographic variable.
CI, confidence interval; OR, odds ratio. Non-linear associations in panels G, H, M, and N were statistically significant at the α=0.05 level.



Supplemental Figure 5.12. Association between sociodemographic variables and odds of being located within 2km versus 5-10km from an oil and gas production high methane emitter. Includes n = 177 exposed and n = 1382 unexposed block groups. Black lines are odds ratios and grey areas represent the 95% confidence intervals. Results from a generalized additive mixed model with a logit link and a random intercept for county adjusted for block group-level percent individuals of non-Hispanic Native American, Asian, and Black race/ethnicity, and percent individuals of Hispanic race/ethnicity, percent individuals living below the federal poverty threshold, percent renters, percent limited English speaking households, percent voter turnout, percent uninsured individuals, and population density. Boxplot displayed along the x-axis shows the number of observations at each level of the respective sociodemographic variable.
CI, confidence interval; OR, odds ratio. Non-linear associations in panels C, D, I, and J were statistically significant at the α=0.05 level.

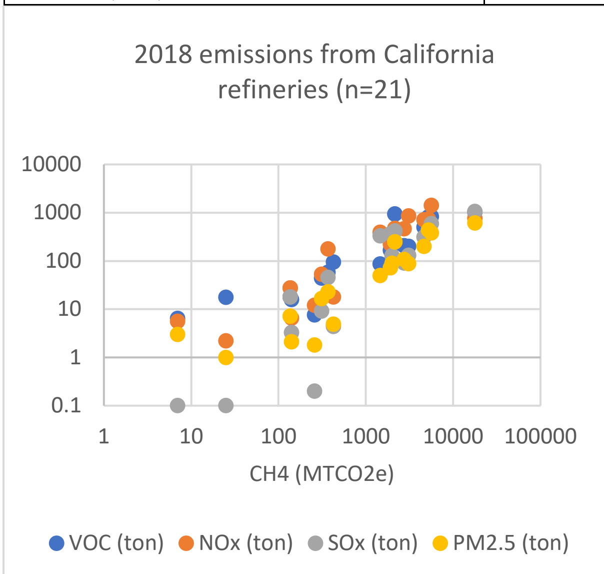


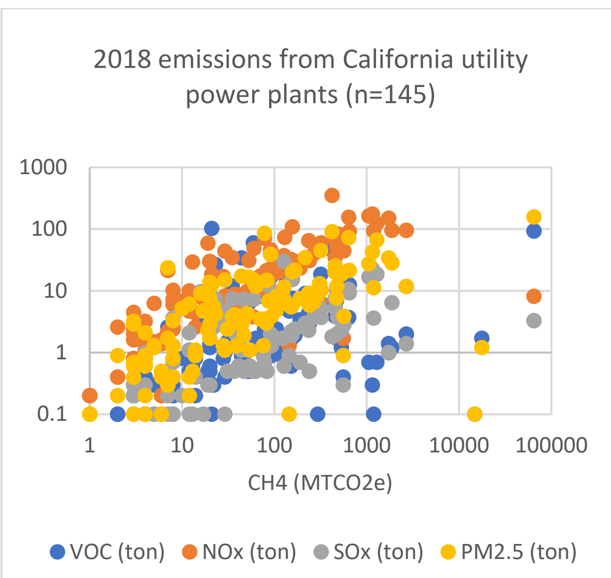
Supplemental Figure 5.13. Association between sociodemographic variables and odds of being located within 2km versus 5-10km from dairy or manure high methane emitter. Includes n = 87 exposed and n = 697 unexposed block groups. Black lines are odds ratios and grey areas represent the 95% confidence intervals. Results from a generalized additive mixed model with a logit link and a random intercept for county adjusted for block group-level percent individuals of non-Hispanic Native American, Asian, and Black race/ethnicity, and percent individuals of Hispanic race/ethnicity, percent individuals living below the federal poverty threshold, percent renters, percent limited English speaking households, percent voter turnout, percent uninsured individuals, and population density. Boxplot displayed along the x-axis shows the number of observations at each level of the respective sociodemographic variable.
CI, confidence interval; OR, odds ratio. The non-linear association in panel J was statistically significant at the α=0.05 level.



Supplemental Figure 5.14. Spearman correlation between 2018 California Air Resources Board Pollution Mapping Tool annual reported CH_4 emissions in MT CO_2e and co-pollutant emissions.

	Refineries (n=21)	Utility power plants (n=145)
VOCs (ton)	0.78	0.47
NOx (ton)	0.80	0.60
SOx (ton)	0.76	0.56
$PM_{2.5}$ (ton)	0.79	0.59





References

Abeysena C, Jayawardana P, Seneviratne R de A. 2010. Effect of psychosocial stress and physical activity on low birthweight: a cohort study. J Obstet Gynaecol Res 36:296–303; doi:10.1111/j.1447-0756.2009.01155.x.

Adetona O, Ozoh OB, Oluseyi T, Uzoegwu Q, Odei J, Lucas M. 2020. An exploratory evaluation of the potential pulmonary, neurological and other health effects of chronic exposure to emissions from municipal solid waste fires at a large dumpsite in Olusosun, Lagos, Nigeria. Environmental Science and Pollution Research 27: 30885–30892.

Adgate JL, Goldstein BD, McKenzie LM. 2014a. Potential Public Health Hazards, Exposures and Health Effects from Unconventional Natural Gas Development. Environ Sci Technol 48:8307–8320; doi:10.1021/es404621d.

Adgate JL, Goldstein BD, McKenzie LM. 2014b. Potential Public Health Hazards, Exposures and Health Effects from Unconventional Natural Gas Development. Environ Sci Technol 48:8307–8320; doi:10.1021/es404621d.

Adgate JL, Goldstein BD, McKenzie LM. 2014c. Potential public health hazards, exposures and health effects from unconventional natural gas development. Environmental science & technology 48: 8307–8320.

Ahmadov R, McKeen S, Trainer M, Banta R, Brewer A, Brown S, et al. 2015. Understanding high wintertime ozone pollution events in an oil- and natural gas-producing region of the western US. Atmospheric Chemistry and Physics 15:411–429; doi:10.5194/acp-15-411-2015.

Alexander GR, Himes JH, Kaufman RB, Mor J, Kogan M. 1996. A United States national reference for fetal growth. Obstet Gynecol 87:163–168; doi:10.1016/0029-7844(95)00386-X.

Alexander GR, Kotelchuck M. 1996. Quantifying the adequacy of prenatal care: a comparison of indices. Public Health Rep 111: 408–419.

Allshouse WB, McKenzie LM, Barton K, Brindley S, Adgate JL. 2019. Community Noise and Air Pollution Exposure During the Development of a Multi-Well Oil and Gas Pad. Environ Sci Technol 53:7126–7135; doi:10.1021/acs.est.9b00052.

Alvarez RA, Zavala-Araiza D, Lyon DR, Allen DT, Barkley ZR, Brandt AR, et al. 2018. Assessment of methane emissions from the U.S. oil and gas supply chain. Science 361:186–188; doi:10.1126/science.aar7204.

Arroyo V, Díaz J, Ortiz C, Carmona R, Sáez M, Linares C. 2016. Short term effect of air pollution, noise and heat waves on preterm births in Madrid (Spain). Environmental Research 145:162–168; doi:10.1016/j.envres.2015.11.034.

Arruti A, Fernández-Olmo I, Irabien A. 2011. Regional evaluation of particulate matter composition in an Atlantic coastal area (Cantabria region, northern Spain): Spatial variations in different urban and rural environments. Atmospheric Research 101:280–293; doi:10.1016/j.atmosres.2011.03.001.

Balazs C, Morello-Frosch R, Hubbard A, Ray I. 2011. Social Disparities in Nitrate-Contaminated Drinking Water in California's San Joaquin Valley. Environmental Health Perspectives 119:1272–1278; doi:10.1289/ehp.1002878.

Balazs Carolina, Morello-Frosch Rachel, Hubbard Alan, Ray Isha. 2011. Social Disparities in Nitrate-Contaminated Drinking Water in California's San Joaquin Valley. Environmental Health Perspectives 119:1272–1278; doi:10.1289/ehp.1002878.

Balazs CL, Morello-Frosch R, Hubbard AE, Ray I. 2012a. Environmental justice implications of arsenic contamination in California's San Joaquin Valley: a cross-sectional, cluster-design examining exposure and compliance in community drinking water systems. Environmental Health 11: 84.

Balazs CL, Morello-Frosch R, Hubbard AE, Ray I. 2012b. Environmental justice implications of arsenic contamination in California's San Joaquin Valley: a cross-sectional, cluster-design examining exposure and compliance in community drinking water systems. Environmental Health 11: 84.

Balazs CL, Ray I. 2014. The Drinking Water Disparities Framework: On the Origins and Persistence of Inequities in Exposure. Am J Public Health 104:603–611; doi:10.2105/AJPH.2013.301664.

Basner M, Müller U, Elmenhorst E-M. 2011. Single and Combined Effects of Air, Road, and Rail Traffic Noise on Sleep and Recuperation. Sleep 34:11–23; doi:10.1093/sleep/34.1.11.

Basu R, Harris M, Sie L, Malig B, Broadwin R, Green R. 2014. Effects of fine particulate matter and its constituents on low birth weight among full-term infants in California. Environmental Research 128:42–51; doi:10.1016/j.envres.2013.10.008.

Bell ML, Banerjee G, Pereira G. 2018. Residential mobility of pregnant women and implications for assessment of spatially-varying environmental exposures. J Expo Sci Environ Epidemiol 28:470–480; doi:10.1038/s41370-018-0026-0.

Bell ML, Belanger K. 2012. Review of research on residential mobility during pregnancy: consequences for assessment of prenatal environmental exposures. Journal of Exposure Science and Environmental Epidemiology 22:429–438; doi:10.1038/jes.2012.42.

Berg J, Ramadan NM. 2006. Societal burden of the headache. The headaches 35–42.

Bhutta AT, Cleves MA, Casey PH, Cradock MM, Anand KJS. 2002. Cognitive and Behavioral Outcomes of School-Aged Children Who Were Born PretermA Meta-analysis. JAMA 288:728–737; doi:10.1001/jama.288.6.728.

Bivand R, Pebesma EJ, Gómez-Rubio V. 2013. *Applied spatial data analysis with R*. second edition. Springer:New York Heidelberg Dordrecht London.

Bivand RS Pebesma, Edzer J, Gomez-Rubio, Virgilio. 2008. *Applied Spatial Data Analysis with R*. Springer:New York, NY.

Blair BD, Brindley S, Dinkeloo E, McKenzie LM, Adgate JL. 2018. Residential noise from nearby oil and gas well construction and drilling. J Expo Sci Environ Epidemiol 28:538–547; doi:10.1038/s41370-018-0039-8.

Blanchard O, Deguen S, Kihal-Talantikite W, François R, Zmirou-Navier D. 2018. Does residential mobility during pregnancy induce exposure misclassification for air pollution? Environmental Health 17:72; doi:10.1186/s12940-018-0416-8.

Bolden AL, Kwiatkowski CF, Colborn T. 2015. New Look at BTEX: Are Ambient Levels a Problem? Environ Sci Technol 49:5261–5276; doi:10.1021/es505316f.

Bonavista Energy Corporation. 2018. News Release: Bonavista Energy Corporation Replaces 189% of 2017 Production with the Addition of 49.8 MMboe of Proved Plus Probable Reserves. Available: https://mma.prnewswire.com/media/636247/Bonavista_Energy_Corporation_Bonavista_Energy_Corporation_Replac.pdf?p=original [accessed 16 July 2018].

Boothroyd IM, Almond S, Qassim SM, Worrall F, Davies RJ. 2016. Fugitive emissions of methane from abandoned, decommissioned oil and gas wells. Science of The Total Environment 547:461–469; doi:10.1016/j.scitotenv.2015.12.096.

Borkum JM. 2016. Migraine Triggers and Oxidative Stress: A Narrative Review and Synthesis. Headache 56:12–35; doi:10.1111/head.12725.

Boyce JK, Pastor M. 2013. Clearing the air: incorporating air quality and environmental justice into climate policy. Climatic Change 120:801–814; doi:10.1007/s10584-013-0832-2.

Boyle MD, Soneja S, Quirós-Alcalá L, Dalemarre L, Sapkota AR, Sangaramoorthy T, et al. 2017. A pilot study to assess residential noise exposure near natural gas compressor stations. PLOS ONE 12:e0174310; doi:10.1371/journal.pone.0174310.

Brandt AR, Heath GA, Cooley D. 2016. Methane Leaks from Natural Gas Systems Follow Extreme Distributions. Environ Sci Technol 50:12512–12520; doi:10.1021/acs.est.6b04303.

Brantley HL, Thoma ED, Eisele AP. 2015a. Assessment of volatile organic compound and hazardous air pollutant emissions from oil and natural gas well pads using mobile remote and on-site direct measurements. Journal of the Air & Waste Management Association 65:1072–1082; doi:10.1080/10962247.2015.1056888.

Brantley HL, Thoma ED, Eisele AP. 2015b. Assessment of volatile organic compound and hazardous air pollutant emissions from oil and natural gas well pads using mobile remote and on-site direct measurements. Journal of the Air & Waste Management Association 65:1072–1082; doi:10.1080/10962247.2015.1056888.

Briggs D, Abellan JJ, Fecht D. 2008. Environmental inequity in England: Small area associations between socio-economic status and environmental pollution. Social Science & Medicine 67:1612–1629; doi:10.1016/j.socscimed.2008.06.040.

Brochu PJ, Yanosky JD, Paciorek CJ, Schwartz J, Chen JT, Herrick RF, et al. 2011. Particulate Air Pollution and Socioeconomic Position in Rural and Urban Areas of the Northeastern United States. Am J Public Health 101:S224–S230; doi:10.2105/AJPH.2011.300232.

Brook JR, Burnett RT, Dann TF, Cakmak S, Goldberg MS, Fan X, et al. 2007. Further interpretation of the acute effect of nitrogen dioxide observed in Canadian time-series studies. Journal of Exposure Science & Environmental Epidemiology 17:S36–S44; doi:10.1038/sj.jes.7500626.

Brown TH. 2016. Diverging Fortunes: Racial/Ethnic Inequality in Wealth Trajectories in Middle and Late Life. Race Soc Probl 8:29–41; doi:10.1007/s12552-016-9160-2.

Bruckner TA, Catalano R. 2018. Selection in utero and population health: Theory and typology of research. SSM Popul Health 5:101–113; doi:10.1016/j.ssmph.2018.05.010.

Brulle RJ, Pellow DN. 2006. ENVIRONMENTAL JUSTICE: Human Health and Environmental Inequalities. Annu Rev Public Health 27:103–124; doi:10.1146/annurev.publhealth.27.021405.102124.

Bryant AS, Worjoloh A, Caughey AB, Washington AE. 2010. Racial/ethnic disparities in obstetric outcomes and care: prevalence and determinants. American Journal of Obstetrics and Gynecology 202:335–343; doi:10.1016/j.ajog.2009.10.864.

Buckley JP, Doherty BT, Keil AP, Engel. 2017. Statistical Approaches for Estimating Sex-Specific Effects in Endocrine Disruptors Research. Environmental Health Perspectives 125:067013; doi:10.1289/EHP334.

Bullard RD. 1983. Solid Waste Sites and the Black Houston Community. Sociological Inquiry 53:273–288; doi:10.1111/j.1475-682X.1983.tb00037.x.

Bullard RD, Mohai P, Saha R, Wright B. 2007. Toxic wastes and race at twenty: 1987-2007. 175.

Buonocore J, Reka S, Yang D, Chang C, Roy A, Thompson T, et al. 2021. Air Quality and Health Benefits of Methane Policies in the United States Oil & Gas Sector.; doi:10.21203/rs.3.rs-199686/v1.

Burger JL, Lovestead TM, Bruno TJ. 2016. Composition of the C6+ Fraction of Natural Gas by Multiple Porous Layer Open Tubular Capillaries Maintained at Low Temperatures. Energy Fuels 30:2119–2126; doi:10.1021/acs.energyfuels.6b00043.

Burton TG, Rifai HS, Hildenbrand ZL, Carlton Jr DD, Fontenot BE, Schug KA. 2016. Elucidating hydraulic fracturing impacts on groundwater quality using a regional geospatial statistical modeling approach. Science of The Total Environment 545–546:114–126; doi:10.1016/j.scitotenv.2015.12.084.

CA DOGGR (currently CA Geologic Energy Management Division). Online Data Oil & Gas - Online Data. Available: http://www.conservation.ca.gov/dog/Online_Data/Pages/Index.aspx [accessed 23 March 2017].

CA DOGGR (currently CA Geologic Energy Management Division). 2019. WST Disclosure. Available: https://www.conservation.ca.gov/calgem/Pages/WSTDisclosureSearchDisclaimer.aspx [accessed 10 May 2019].

CalGEM (CA Geologic Energy Management Division). 2019a. GIS Mapping - All Wells Shapefile. Available: https://www.conservation.ca.gov/calgem/maps/Pages/GISMapping2.aspx [accessed 28 May 2020].

CalGEM (CA Geologic Energy Management Division). 2015. Legislative Reports: Well Stimulation Treatment Annual Report (SB 4). Available: https://www.conservation.ca.gov/calgem/pubs_stats/Pages/legislative_reports.aspx#wst-annual-report [accessed 4 June 2020].

CalGEM (CA Geologic Energy Management Division). 2019b. Well Stimulation Treatment Disclosure Database. Available: ftp://ftp.consrv.ca.gov/pub/oil/Well_Stimulation_Treatment_Disclosures/Historical%20Disclosure%20Data/ [accessed 10 May 2019].

CalGEM (CA Geologic Energy Management Division). 2019c. WellSTAR EPA Aquifer Exemption Review: Disposal Wells, Enhanced Oil Recovery Wells. Available: https://data.ca.gov/organization/california-department-of-conservation?q=epa+aquifer&sort=score+desc%2C+metadata_modified+desc [accessed 12 November 2019].

California AB1328 | TrackBill. Available: https://trackbill.com/bill/california-assembly-bill-1328-oil-and-gas-notice-of-intention-to-abandon-well-study-of-fugitive-emissions-from-idle-idle-deserted-and-abandoned-wells/1700789/ [accessed 24 October 2019].

California Air Resources Board. 2020. Landfill Activities. Available: https://ww2.arb.ca.gov/sites/default/files/classic/cc/landfills/landfills.htm.

California Air Resources Board. 2019. Refiner Emergency Air Monitoring Assessment Report Objective 2: Evaluation of Air Monitoring Capabilities, Gaps, and Potential Enhancements. Available: https://ww3.arb.ca.gov/fuels/carefinery/crseam/o2reamarfinal.pdf?_ga=2.149873781.258059090.1593094388-358911420.1591295298.

CARB (CA Air Resources Board). 2017. CARB Pollution Mapping Tool. Available: https://ww3.arb.ca.gov/ei/tools/pollution_map/ [accessed 17 September 2019].

CARB (CA Air Resources Board). 2014. Geographical Information System (GIS) Library. Available: https://www.arb.ca.gov/ei/gislib/gislib.htm [accessed 18 September 2018].

Caron-Beaudoin É, Valter N, Chevrier J, Ayotte P, Frohlich K, Verner M-A. 2018. Gestational exposure to volatile organic compounds (VOCs) in Northeastern British Columbia, Canada: A pilot study. Environment International 110:131–138; doi:10.1016/j.envint.2017.10.022.

Caron-Beaudoin É, Whitworth KW, Bosson-Rieutort D, Wendling G, Liu S, Verner M-A. 2021. Density and proximity to hydraulic fracturing wells and birth outcomes in Northeastern British Columbia, Canada. J Expo Sci Environ Epidemiol 31:53–61; doi:10.1038/s41370-020-0245-z.

Casey JA, Kim BF, Larsen J, Price LB, Nachman KE. 2015a. Industrial Food Animal Production and Community Health. Curr Envir Health Rpt 2:259–271; doi:10.1007/s40572-015-0061-0.

Casey JA, Kim BF, Larsen J, Price LB, Nachman KE. 2015b. Industrial Food Animal Production and Community Health. Curr Environ Health Rep 2:259–271; doi:10.1007/s40572-015-0061-0.

Casey JA, Savitz DA, Rasmussen SG, Ogburn EL, Pollak J, Mercer DG, et al. 2015c. Unconventional Natural Gas Development and Birth Outcomes in Pennsylvania, USA: Epidemiology 1; doi:10.1097/EDE.0000000000000387.

CCST (CA Council on Science and Technology). 2015. Vol 1. Appendix M: Integrated hydraulic fracturing data set regarding occurrence, location, date, and depth cover sheet. Available: https://ccst.us/wp-content/uploads/Appendix_M_cover.pdf [accessed 29 October 2015].

CDPH (CA Deparment of Public Health). 2015. Maternal Smoking Infographic 2015.

Chabriat H, Danchot J, Michel P, Joire J, Henry P. 1999. Precipitating factors of headache. A prospective study in a national control-matched survey in migraineurs and nonmigraineurs. Headache: The journal of head and face pain 39: 335–338.

Chang M, Park H, Ha M, Hong Y-C, Lim Y-H, Kim Y, et al. 2017. The effect of prenatal TVOC exposure on birth and infantile weight: the Mothers and Children's Environmental Health study. Pediatr Res 82:423–428; doi:10.1038/pr.2017.55.

Charles A. 2013. The evolution of a migraine attack–a review of recent evidence. Headache: The Journal of Head and Face Pain 53: 413–419.

Chavis BF, Lee C. 1987. Toxic Wastes and Race in the United States. 69.

Chen C-C, Tsai S-S, Yang C-Y. 2015. Association between fine particulate air pollution and daily clinic visits for migraine in a subtropical city: Taipei, Taiwan. International journal of environmental research and public health 12: 4697–4708.

Chen L, Bell EM, Caton AR, Druschel CM, Lin S. 2010. Residential mobility during pregnancy and the potential for ambient air pollution exposure misclassification. Environmental Research 110:162–168; doi:10.1016/j.envres.2009.11.001.

Chiu H-F, Weng Y-H, Chiu Y-W, Yang C-Y. 2015. Air Pollution and Daily Clinic Visits for Headache in a Subtropical City: Taipei, Taiwan. Int J Environ Res Public Health 12:2277–2288; doi:10.3390/ijerph120202277.

Clark LP, Millet DB, Marshall JD. 2014. National Patterns in Environmental Injustice and Inequality: Outdoor NO2 Air Pollution in the United States. PLoS ONE 9:e94431; doi:10.1371/journal.pone.0094431.

Clough E, Bell D. 2016a. Just fracking: a distributive environmental justice analysis of unconventional gas development in Pennsylvania, USA. Environ Res Lett 11:025001; doi:10.1088/1748-9326/11/2/025001.

Clough E, Bell D. 2016b. Just fracking: a distributive environmental justice analysis of unconventional gas development in Pennsylvania, USA. Environ Res Lett 11:025001; doi:10.1088/1748-9326/11/2/025001.

Colborn T, Schultz K, Herrick L, Kwiatkowski C. 2014. An Exploratory Study of Air Quality Near Natural Gas Operations. Human and Ecological Risk Assessment: An International Journal 20:86–105; doi:10.1080/10807039.2012.749447.

Collett JL, Colorado State University. 2016. Characterizing Emissions from Natural Gas Drilling and Well Completion Operations in Garfield County, CO.

Colmer J, Hardman I, Shimshack J, Voorheis J. 2020. Disparities in PM2. 5 air pollution in the United States. Science 369: 575–578.

Community Air Protection Blueprint | California Air Resources Board. Available: https://ww2.arb.ca.gov/capp-blueprint [accessed 28 June 2021].

Cooke L, Rose M, Becker W. 2000. Chinook winds and migraine headache. Neurology 54: 302–302.

Costa LG, Cole TB, Coburn J, Chang Y-C, Dao K, Roqué PJ. 2017. Neurotoxicity of traffic-related air pollution. Neurotoxicology 59: 133–139.

Cumming DG. 2018. Black Gold, White Power: Mapping Oil, Real Estate, and Racial Segregation in the Los Angeles Basin, 1900-1939. Engaging Science, Technology, and Society 4:85; doi:10.17351/ests2018.212.

Currie J, Greenstone M, Meckel K. 2017. Hydraulic fracturing and infant health: New evidence from Pennsylvania. Science Advances 3:e1603021; doi:10.1126/sciadv.1603021.

Cushing L, Faust J, August LM, Cendak R, Wieland W, Alexeeff G. 2015. Racial/Ethnic Disparities in Cumulative Environmental Health Impacts in California: Evidence From a Statewide Environmental Justice Screening Tool (CalEnviroScreen 1.1). Am J Public Health 105:2341–2348; doi:10.2105/AJPH.2015.302643.

Cushing LJ, Vavra-Musser K, Chau K, Franklin M, Johnston JE. 2020. Flaring from Unconventional Oil and Gas Development and Birth Outcomes in the Eagle Ford Shale in South Texas. Environ Health Perspect 128:077003; doi:10.1289/EHP6394.

CVRWQCB (Central Valley Regional Water Quality Control Board). 2019a. Central Valley Regional Water Quality Control Board Wastewater Disposal Pond List.

CVRWQCB (Central Valley Regional Water Quality Control Board). 2019b. Press Release: Cease and Desist Order Issued for Oil Field Wastewater Disposal Facility in Kern County to Ensure Greater Groundwater Protections.

Dadvand P, Basagaña X, Figueras F, Martinez D, Beelen R, Cirach M, et al. 2014. Air Pollution and Preterm Premature Rupture of Membranes: A Spatiotemporal Analysis. Am J Epidemiol 179:200–207; doi:10.1093/aje/kwt240.

Dadvand P, Parker J, Bell ML, Bonzini M, Brauer M, Darrow LA, et al. 2013. Maternal Exposure to Particulate Air Pollution and Term Birth Weight: A Multi-Country Evaluation of Effect and Heterogeneity.

Davidson CI, Phalen RF, Solomon PA. 2005. Airborne particulate matter and human health: a review. Aerosol Science and Technology 39: 737–749.

De Lissovoy G, Lazarus SS. 1994. The economic cost of migraine. Present state of knowledge. Neurology 44: S56.

de Mesnard L. 2013. Pollution models and inverse distance weighting: Some critical remarks. Computers & Geosciences 52:459–469; doi:10.1016/j.cageo.2012.11.002.

Deziel NC. 2021. Invited Perspective: Oil and Gas Development and Adverse Birth Outcomes: What More Do We Need to Know? Environmental Health Perspectives 129:071301; doi:10.1289/EHP9715.

Deziel NC, Brokovich E, Grotto I, Clark CJ, Barnett-Itzhaki Z, Broday D, et al. 2020. Unconventional oil and gas development and health outcomes: A scoping review of the epidemiological research. Environmental Research 182:109124; doi:10.1016/j.envres.2020.109124.

DiGiulio DC, Shonkoff SBC. 2019. Risk Posed to Groundwater Resources by the Disposal of Produced Water into Unlined Produced Water Ponds in California.

Dominguez TP, Dunkel-Schetter C, Glynn LM, Hobel C, Sandman CA. 2008. Racial differences in birth outcomes: the role of general, pregnancy, and racism stress. Health Psychol 27:194–203; doi:10.1037/0278-6133.27.2.194.

Donham KJ, Wing S, Osterberg D, Flora JL, Hodne C, Thu KM, et al. 2007a. Community Health and Socioeconomic Issues Surrounding Concentrated Animal Feeding Operations. Environ Health Perspect 115:317–320; doi:10.1289/ehp.8836.

Donham KJ, Wing S, Osterberg D, Flora JL, Hodne C, Thu KM, et al. 2007b. Community Health and Socioeconomic Issues Surrounding Concentrated Animal Feeding Operations. Environ Health Perspect 115:317–320; doi:10.1289/ehp.8836.

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. PNAS 112:13184–13189; doi:10.1073/pnas.1511474112.

Duren RM, Thorpe AK, Foster KT, Rafiq T, Hopkins FM, Yadav V, et al. 2019a. California's methane super-emitters. Nature 575:180–184; doi:10.1038/s41586-019-1720-3.

Duren RM, Thorpe AK, Foster KT, Rafiq T, Hopkins FM, Yadav V, et al. 2019b. California's methane super-emitters. Nature 575:180–184; doi:10.1038/s41586-019-1720-3.

Eapi GR, Sabnis MS, Sattler ML. 2014. Mobile measurement of methane and hydrogen sulfide at natural gas production site fence lines in the Texas Barnett Shale. Journal of the Air & Waste Management Association 64:927–944; doi:10.1080/10962247.2014.907098.

Ebisu K, Bell ML. 2012. Airborne PM 2.5 Chemical Components and Low Birth Weight in the Northeastern and Mid-Atlantic Regions of the United States. Environmental Health Perspectives 120: 1746–1752.

Eilerman SJ, Peischl J, Neuman JA, Ryerson TB, Aikin KC, Holloway MW, et al. 2016. Characterization of Ammonia, Methane, and Nitrous Oxide Emissions from Concentrated Animal Feeding Operations in Northeastern Colorado. Environ Sci Technol 50:10885–10893; doi:10.1021/acs.est.6b02851.

Eisele AP, Mukerjee S, Smith LA, Thoma ED, Whitaker DA, Oliver KD, et al. 2016. Volatile organic compounds at two oil and natural gas production well pads in Colorado and Texas using passive samplers. J Air Waste Manag Assoc 66:412–419; doi:10.1080/10962247.2016.1141808.

Elliott EG, Ettinger AS, Leaderer BP, Bracken MB, Deziel NC. 2017. A systematic evaluation of chemicals in hydraulic-fracturing fluids and wastewater for reproductive and developmental toxicity. J Expos Sci Environ Epidemiol 27:90–99; doi:10.1038/jes.2015.81.

Energy Information Agency. 2020. California State Energy Profile. Available: https://www.eia.gov/state/print.php?sid=CA [accessed 17 July 2020].

Eross E, Dodick D, Eross M. 2007. The Sinus, Allergy and Migraine Study (SAMS) CME. Headache: The Journal of Head and Face Pain 47: 213–224.

Esswein EJ, Snawder J, King B, Breitenstein M, Alexander-Scott M, Kiefer M. 2014. Evaluation of Some Potential Chemical Exposure Risks During Flowback Operations in Unconventional Oil and Gas Extraction: Preliminary Results. Journal of Occupational and Environmental Hygiene 11:D174–D184; doi:10.1080/15459624.2014.933960.

Feng S, Gao D, Liao F, Zhou F, Wang X. 2016. The health effects of ambient PM2.5 and potential mechanisms. Ecotoxicology and Environmental Safety 128:67–74; doi:10.1016/j.ecoenv.2016.01.030.

Fernandez-Bou AS, Ortiz-Partida JP, Dobbin KB, Flores-Landeros H, Bernacchi LA, Medellín-Azuara J. 2021. Underrepresented, understudied, underserved: Gaps and opportunities for advancing justice in disadvantaged communities. Environmental Science & Policy 122:92–100; doi:10.1016/j.envsci.2021.04.014.

Ferrar KJ, Kriesky J, Christen CL, Marshall LP, Malone SL, Sharma RK, 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. International Journal of Occupational and Environmental Health 19:104–112; doi:10.1179/2049396713Y.0000000024.

Ferrari MD. 1998. The economic burden of migraine to society. Pharmacoeconomics 13: 667–676.

Finkel ML, Law A. 2011. The Rush to Drill for Natural Gas: A Public Health Cautionary Tale. Am J Public Health 101:784–785; doi:10.2105/AJPH.2010.300089.

Finkelstein MM, Jerrett M, DeLuca P, Finkelstein N, Verma DK, Chapman K, et al. 2003. Relation between income, air pollution and mortality: a cohort study. CMAJ 169: 397–402.

Franklin M, Chau K, Cushing LJ, Johnston JE. 2019. Characterizing Flaring from Unconventional Oil and Gas Operations in South Texas Using Satellite Observations. Environ Sci Technol 53:2220–2228; doi:10.1021/acs.est.8b05355.

Friedman DI, De Ver Dye T. 2009. Migraine and the environment. Headache: The Journal of Head and Face Pain 49: 941–952.

Garcia C. 2017. CA Assembly Bill 617. Nonvehicular air pollution: criteria air pollutants and toxic air contaminants.

Garcia-Gonzales DA, Popoola O, Bright VB, Paulson SE, Wang Y, Jones RL, et al. 2019a. Associations among particulate matter, hazardous air pollutants and methane emissions from the Aliso Canyon natural gas storage facility during the 2015 blowout. Environment International 132:104855; doi:10.1016/j.envint.2019.05.049.

Garcia-Gonzales DA, Shonkoff SBC, Hays J, Jerrett M. 2019b. Hazardous Air Pollutants Associated with Upstream Oil and Natural Gas Development: A Critical Synthesis of Current Peer-Reviewed Literature. Annu Rev Public Health 40:283–304; doi:10.1146/annurev-publhealth-040218-043715.

Garcia-Gonzales DA, Shonkoff SBC, Hays J, Jerrett M. 2019c. Hazardous Air Pollutants Associated with Upstream Oil and Natural Gas Development: A Critical Synthesis of Current Peer-Reviewed Literature. Annual Review of Public Health 40:283–304; doi:10.1146/annurev-publhealth-040218-043715.

Gehring U, Tamburic L, Sbihi H, Davies HW, Brauer M. 2014. Impact of Noise and Air Pollution on Pregnancy Outcomes: Epidemiology 25:351–358; doi:10.1097/EDE.0000000000000073.

Giffin N, Ruggiero L, Lipton RB, Silberstein S, Tvedskov J, Olesen J, et al. 2003. Premonitory symptoms in migraine: an electronic diary study. Neurology 60: 935–940.

Gilligan AM, Foster SA, Sainski-Nguyen A, Sedgley R, Smith D, Morrow P. 2018. Direct and indirect costs among United States commercially insured employees with migraine. Journal of Occupational and Environmental Medicine 60: 1120–1127.

Gilman JB, Lerner BM, Kuster WC, de Gouw JA. 2013. Source Signature of Volatile Organic Compounds from Oil and Natural Gas Operations in Northeastern Colorado. Environ Sci Technol 47:1297–1305; doi:10.1021/es304119a.

Goddard JJ. 2019. Measuring Drinking Water Affordability and Sustainability. UC Berkeley.

Goetz JD, Floerchinger C, Fortner EC, Wormhoudt J, Massoli P, Knighton WB, et al. 2015. Atmospheric Emission Characterization of Marcellus Shale Natural Gas Development Sites. Environ Sci Technol 49:7012–7020; doi:10.1021/acs.est.5b00452.

Goin DE, Casey JA, Kioumourtzoglou M-A, Cushing LJ, Morello-Frosch R. 2021. Environmental hazards, social inequality, and fetal loss: Implications of live-birth bias for estimation of disparities in birth outcomes. Environmental Epidemiology 5: e131.

Goldberg DMS, Al-Homsi N, Goulet L, Riberdy H. 1995. Incidence of Cancer among Persons Living Near a Municipal Solid Waste Landfill Site in Montreal, Québec. Archives of Environmental Health: An International Journal 50:416–424; doi:10.1080/00039896.1995.9935977.

Goldenberg RL, Culhane JF, Iams JD, Romero R. 2008. Epidemiology and causes of preterm birth. The Lancet 371:75–84; doi:10.1016/S0140-6736(08)60074-4.

Gonzalez DJX, Sherris AR, Yang W, Stevenson DK, Padula AM, Baiocchi M, et al. 2020a. Oil and gas production and spontaneous preterm birth in the San Joaquin Valley, CA: A case–control study. Environmental Epidemiology 4:e099; doi:10.1097/EE9.0000000000000099.

Gonzalez DJX, Sherris AR, Yang W, Stevenson DK, Padula AM, Baiocchi M, et al. 2020b. Oil and gas production and spontaneous preterm birth in the San Joaquin Valley, CA: A case–control study. Environmental Epidemiology 4:e099; doi:10.1097/EE9.0000000000000099.

Gray SC, Edwards SE, Schultz BD, Miranda ML. 2014. Assessing the impact of race, social factors and air pollution on birth outcomes: a population-based study. Environmental Health 13:4; doi:10.1186/1476-069X-13-4.

Grineski SE, Collins TW, Morales DX. 2017. Asian Americans and disproportionate exposure to carcinogenic hazardous air pollutants: A national study. Social Science & Medicine 185:71–80; doi:10.1016/j.socscimed.2017.05.042.

Gross SA, Avens HJ, Banducci AM, Sahmel J, Panko JM, Tvermoes BE. 2013a. Analysis of BTEX groundwater concentrations from surface spills associated with hydraulic fracturing operations. Journal of the Air & Waste Management Association 63:424–432; doi:10.1080/10962247.2012.759166.

Gross SA, Avens HJ, Banducci AM, Sahmel J, Panko JM, Tvermoes BE. 2013b. Analysis of BTEX groundwater concentrations from surface spills associated with hydraulic fracturing operations. J Air Waste Manag Assoc 63: 424–432.

Ground Water Protection Council. 2011. State Oil and Gas Agency Groundwater InvestigationsAnd Their Role in Advancing Regulatory Reforms, A Two-State Review: Ohio and Texas.

Hack M, Klein NK, Taylor HG. 1995. Long-Term Developmental Outcomes of Low Birth Weight Infants. The Future of Children 5:176–196; doi:10.2307/1602514.

Halliday HS, Thompson AM, Wisthaler A, Blake DR, Hornbrook RS, Mikoviny T, et al. 2016. Atmospheric benzene observations from oil and gas production in the Denver-Julesburg Basin in July and August 2014. Journal of Geophysical Research: Atmospheres 121:11,055-11,074; doi:10.1002/2016JD025327.

Hanak E, Escriva-Bou A, Gray B, Green S, Harter T, Jezdimirovic J, et al. 2019. Water and the Future of the San Joaquin Valley.

Hays J, McCawley M, Shonkoff SBC. 2017a. Public health implications of environmental noise associated with unconventional oil and gas development. Science of The Total Environment 580:448–456; doi:10.1016/j.scitotenv.2016.11.118.

Hays J, McCawley M, Shonkoff SBC. 2017b. Public health implications of environmental noise associated with unconventional oil and gas development. Sci Total Environ 580:448–456; doi:10.1016/j.scitotenv.2016.11.118.

Hecobian A, Clements AL, Shonkwiler K, Zhou Y, MacDonald L, Hilliard N, et al. 2019. Air toxics and other volatile organic compound emissions from unconventional oil and gas development. Environ Sci Technol Lett; doi:10.1021/acs.estlett.9b00591.

Helmig D, Thompson CR, Evans J, Boylan P, Hueber J, Park J-H. 2014. Highly Elevated Atmospheric Levels of Volatile Organic Compounds in the Uintah Basin, Utah. Environ Sci Technol 48:4707–4715; doi:10.1021/es405046r.

Henry P, Auray J, Gaudin A, Dartigues J, Duru G, Lantéri–Minet M, et al. 2002. Prevalence and clinical characteristics of migraine in France. Neurology 59: 232–237.

Hildenbrand ZL, Carlton DD, Fontenot BE, Meik JM, Walton JL, Taylor JT, et al. 2015. A Comprehensive Analysis of Groundwater Quality in The Barnett Shale Region. Environ Sci Technol 49:8254–8262; doi:10.1021/acs.est.5b01526.

Hildenbrand ZL, Carlton Jr. DD, Fontenot BE, Meik JM, Walton JL, Thacker JB, et al. 2016. Temporal variation in groundwater quality in the Permian Basin of Texas, a region of increasing unconventional oil and gas development. Science of The Total Environment 562:906–913; doi:10.1016/j.scitotenv.2016.04.144.

Hill EL. 2018. Shale gas development and infant health: Evidence from Pennsylvania. Journal of Health Economics 61:134–150; doi:10.1016/j.jhealeco.2018.07.004.

Hirsch JK, Bryant Smalley K, Selby-Nelson EM, Hamel-Lambert JM, Rosmann MR, Barnes TA, et al. 2018. Psychosocial Impact of Fracking: a Review of the Literature on the Mental Health Consequences of Hydraulic Fracturing. Int J Ment Health Addiction 16:1–15; doi:10.1007/s11469-017-9792-5.

Hodgson S, Lurz PWW, Shirley MDF, Bythell M, Rankin J. 2015. Exposure misclassification due to residential mobility during pregnancy. International Journal of Hygiene and Environmental Health 218:414–421; doi:10.1016/j.ijheh.2015.03.007.

Hong S-Y, Wan L, Lin H-J, Lin C-L, Wei C-C. 2020. Long-Term Ambient Air Pollutant Exposure and Risk of Recurrent Headache in Children: A 12-Year Cohort Study. International Journal of Environmental Research and Public Health 17: 9140.

Howarth RW. 2019. Ideas and perspectives: is shale gas a major driver of recent increase in global atmospheric methane? Biogeosciences 16:3033–3046; doi:10.5194/bg-16-3033-2019.

Howarth RW, Ingraffea A, Engelder T. 2011. Should fracking stop? Nature 477:271–275; doi:10.1038/477271a.

Hu XH, Markson LE, Lipton RB, Stewart WF, Berger ML. 1999. Burden of migraine in the United States: disability and economic costs. Archives of internal medicine 159: 813–818.

Hubbard AE, Ahern J, Fleischer NL, Van der Laan M, Satariano SA, Jewell N, et al. 2010. To GEE or Not to GEE: Comparing Population Average and Mixed Models for Estimating the Associations Between Neighborhood Risk Factors and Health. Epidemiology 21: 467–474.

Hyder A, Lee HJ, Ebisu K, Koutrakis P, Belanger K, Bell ML. 2014. PM2.5 exposure and birth outcomes: Use of satellite- and monitor-based data. Epidemiology 25:58–67; doi:10.1097/EDE.0000000000000027.

Ingraffea AR, Wells MT, Santoro RL, Shonkoff SBC. 2014. Assessment and risk analysis of casing and cement impairment in oil and gas wells in Pennsylvania, 2000-2012. Proceedings of the National Academy of Sciences 111:10955–10960; doi:10.1073/pnas.1323422111.

Jackson RB, Lowry ER, Pickle A, Kang M, DiGiulio D, Zhao K. 2015. The Depths of Hydraulic Fracturing and Accompanying Water Use Across the United States. Environ Sci Technol 49:8969–8976; doi:10.1021/acs.est.5b01228.

Jackson RB, Saunois M, Bousquet P, Canadell JG, Poulter B, Stavert AR, et al. 2020. Increasing anthropogenic methane emissions arise equally from agricultural and fossil fuel sources. Environ Res Lett 15:071002; doi:10.1088/1748-9326/ab9ed2.

Jackson SC. 2009. Parallel Pursuit of Near-Term and Long-Term Climate Mitigation. Science 326:526–527; doi:10.1126/science.1177042.

Jalaludin B, Mannes T, Morgan G, Lincoln D, Sheppeard V, Corbett S. 2007. Impact of ambient air pollution on gestational age is modified by season in Sydney, Australia. Environ Health 6:16; doi:10.1186/1476-069X-6-16.

Janitz AE, Dao HD, Campbell JE, Stoner JA, Peck JD. 2019. The association between natural gas well activity and specific congenital anomalies in Oklahoma, 1997–2009. Environment International 122:381–388; doi:10.1016/j.envint.2018.12.011.

Jasechko S, Perrone D. 2017. Hydraulic fracturing near domestic groundwater wells. PNAS 114:13138–13143; doi:10.1073/pnas.1701682114.

Jerrett M, Burnett RT, Pope CA, Ito K, Thurston G, Krewski D, et al. 2009. Long-Term Ozone Exposure and Mortality. New England Journal of Medicine 360:1085–1095; doi:10.1056/NEJMoa0803894.

Jewell NP. 2003. *Statistics for epidemiology*. CRC Press.

Johnston JE, Chau K, Franklin M, Cushing L. 2020a. Environmental Justice Dimensions of Oil and Gas Flaring in South Texas: Disproportionate Exposure among Hispanic communities. Environ Sci Technol; doi:10.1021/acs.est.0c00410.

Johnston JE, Chau K, Franklin M, Cushing L. 2020b. Environmental Justice Dimensions of Oil and Gas Flaring in South Texas: Disproportionate Exposure among Hispanic communities. Environ Sci Technol 54:6289–6298; doi:10.1021/acs.est.0c00410.

Johnston JE, Lim E, Roh H. 2019a. Impact of upstream oil extraction and environmental public health: A review of the evidence. Science of The Total Environment 657:187–199; doi:10.1016/j.scitotenv.2018.11.483.

Johnston JE, Lim E, Roh H. 2019b. Impact of upstream oil extraction and environmental public health: A review of the evidence. Sci Total Environ 657:187–199; doi:10.1016/j.scitotenv.2018.11.483.

Johnston JE, Werder E, Sebastian D. 2016a. Wastewater Disposal Wells, Fracking, and Environmental Injustice in Southern Texas. Am J Public Health 106:550–556; doi:10.2105/AJPH.2015.303000.

Johnston JE, Werder E, Sebastian D. 2016b. Wastewater Disposal Wells, Fracking, and Environmental Injustice in Southern Texas. American Journal of Public Health 106:550–556; doi:10.2105/AJPH.2015.303000.

Kang M, Christian S, Celia MA, Mauzerall DL, Bill M, Miller AR, et al. 2016. Identification and characterization of high methane-emitting abandoned oil and gas wells. PNAS 113:13636–13641; doi:10.1073/pnas.1605913113.

Kelly-Reif K, Wing S. 2016. Urban-rural exploitation: An underappreciated dimension of environmental injustice. Journal of Rural Studies 47:350–358; doi:10.1016/j.jrurstud.2016.03.010.

Kendrick CM, Koonce P, George LA. 2015. Diurnal and seasonal variations of NO, NO 2 and PM 2.5 mass as a function of traffic volumes alongside an urban arterial. Atmospheric Environment 122:133–141; doi:10.1016/j.atmosenv.2015.09.019.

Kim SY, Bechle M, Hankey S, Sheppard L, Szpiro AA, Marshall JD. 2020a. Concentrations of criteria pollutants in the contiguous U.S., 1979 – 2015: Role of prediction model parsimony in integrated empirical geographic regression. PLoS ONE 15; doi:10.1371/journal.pone.0228535.

Kim SY, Bechle M, Hankey S, Sheppard L, Szpiro AA, Marshall JD. 2020b. Concentrations of criteria pollutants in the contiguous U.S., 1979 - 2015: Role of prediction model parsimony in integrated empirical geographic regression. PLoS One 15:e0228535; doi:10.1371/journal.pone.0228535.

Koss A, Yuan B, Warneke C, Gilman JB, Lerner BM, Veres PR, et al. 2017. Observations of VOC emissions and photochemical products over US oil- and gas-producing regions using high-resolution H_3O^+ CIMS (PTR-ToF-MS). Atmospheric Measurement Techniques 10:2941–2968; doi:10.5194/amt-10-2941-2017.

Koss AR, Yuan B, Warneke C, Gilman JB, Lerner BM, Veres PR, et al. Observations of VOC emissions and photochemical products over US oil- and gas-producing regions using high-resolution H3O+ CIMS (PTR-ToF-MS). 30.

Kotelchuck M. 1994. An evaluation of the Kessner adequacy of prenatal care index and a proposed adequacy of prenatal care utilization index. American journal of public health 84: 1414–1420.

Kovats S, Depledge M, Haines A, Fleming LE, Wilkinson P, Shonkoff SB, et al. 2014. The health implications of fracking. The Lancet 383:757–758; doi:10.1016/S0140-6736(13)62700-2.

Krieger N, Chen J, Waterman P, Soobader M, Subramanian S, Carson R. 2003. Choosing area based socioeconomic measures to monitor social inequalities in low birth weight and childhood lead poisoning: The Public Health Disparities Geocoding Project (US). J Epidemiol Community Health 57:186–199; doi:10.1136/jech.57.3.186.

Krieger N, Waterman PD, Spasojevic J, Li W, Maduro G, Van Wye G. 2016. Public health monitoring of privilege and deprivation with the index of concentration at the extremes. American journal of public health 106: 256–263.

Kroepsch AC, Maniloff PT, Adgate JL, McKenzie LM, Dickinson KL. 2019a. Environmental Justice in Unconventional Oil and Natural Gas Drilling and Production: A Critical Review and Research Agenda. Environ Sci Technol 53:6601–6615; doi:10.1021/acs.est.9b00209.

Kroepsch AC, Maniloff PT, Adgate JL, McKenzie LM, Dickinson KL. 2019b. Environmental Justice in Unconventional Oil and Natural Gas Drilling and Production: A Critical Review and Research Agenda. Environ Sci Technol 53:6601–6615; doi:10.1021/acs.est.9b00209.

Kunkler PE, Zhang L, Johnson PL, Oxford GS, Hurley JH. 2018. Induction of chronic migraine phenotypes in a rat model after environmental irritant exposure. Pain 159: 540.

Kunkler PE, Zhang L, Pellman JJ, Oxford GS, Hurley JH. 2015. Sensitization of the trigeminovascular system following environmental irritant exposure. Cephalalgia 35: 1192–1201.

LACDPH (LA County Department of Public Health). 2018. Public Health and Safety Risks of Oil and Gas Facilities in Los Angeles County.

Ladd AE, Edward B. 2002. Corporate Swine and Capitalist Pigs: A Decade of Environmental Injustice and Protest in North Carolina. Social Justice 29: 26–46.

Lan X, Talbot R, Laine P, Torres A, Lefer B, Flynn J. 2015. Atmospheric Mercury in the Barnett Shale Area, Texas: Implications for Emissions from Oil and Gas Processing. Environ Sci Technol 49:10692–10700; doi:10.1021/acs.est.5b02287.

Le Rest K, Pinaud D, Bretagnolle V. 2013. Accounting for spatial autocorrelation from model selection to statistical inference: Application to a national survey of a diurnal raptor. Ecological Informatics 14:17–24; doi:10.1016/j.ecoinf.2012.11.008.

Lee H, Myung W, Cheong H-K, Yi S-M, Hong Y-C, Cho S-I, et al. 2018. Ambient air pollution exposure and risk of migraine: Synergistic effect with high temperature. Environment international 121: 383–391.

Li R, Zhang J, Zhou R, Liu J, Dai Z, Liu D, et al. 2017. Sleep disturbances during pregnancy are associated with cesarean delivery and preterm birth. The Journal of Maternal-Fetal & Neonatal Medicine 30:733–738; doi:10.1080/14767058.2016.1183637.

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

Lipton RB, Stewart WF, Diamond S, Diamond ML, Reed M. 2001a. Prevalence and burden of migraine in the United States: data from the American Migraine Study II. Headache: The Journal of Head and Face Pain 41: 646–657.

Lipton RB, Stewart WF, Scher AI. 2001b. Epidemiology and economic impact of migraine. Current medical research and opinion 17: s4-12.

Llewellyn GT, Dorman F, Westland JL, Yoxtheimer D, Grieve P, Sowers T, et al. 2015. Evaluating a groundwater supply contamination incident attributed to Marcellus Shale gas development. PNAS 112:6325–6330; doi:10.1073/pnas.1420279112.

Lofland J. 2007. Impact of migraine headache in the United States. Adv Stud Pharm 4: 8–10.

Long JCS, Feinstein LC, Bachmann CE, Birkholzer JT, Camarillo MK, Domen JK, et al. 2015a. An Independent Scientific Assessment of Well Stimulation in California Volume II: Potential Environmental Impacts of Hydraulic Fracturing and Acid Stimulations.

Long JCS, Feinstein LC, Birkholzer JT, Foxall W, Houseworth JE, Jordan PD, et al. 2015b. An Independent Scientific Assessment of Well Stimulation in California Volume III: Case Studies of Hydraulic Fracturing and Acid Stimulations in Select Regions: Onshore, Monterey Formation, Los Angeles Basin, and San Joaquin Basin.

Long JCS, Feinstein LC, Birkholzer JT, Jordan PD, Houseworth JE, Dobson PF, et al. 2015c. An Independent Scientific Assessment of Well Stimulation in California Volume I: Well Stimulation Technologies and their Past, Present, and Potential Future Use in California.

Lupo PJ, Symanski E, Chan W, Mitchell LE, Waller DK, Canfield MA, et al. 2010. Differences in exposure assignment between conception and delivery: the impact of maternal mobility. Paediatric and Perinatal Epidemiology 24:200–208; doi:10.1111/j.1365-3016.2010.01096.x.

Ma Z. 2016. Time Series Evaluation of Birth Defects in Areas with and without Unconventional Natural Gas Development. Journal of Epidemiology and Public Health Reviews 1; doi:10.16966/2471-8211.107.

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. Environmental Health 13:82; doi:10.1186/1476-069X-13-82.

Maizlish N. 2016. Technical Documentation: California Health Disadvantage Index (HDI 1.1). 38.

Malin SA, DeMaster KT. 2016. A devil's bargain: Rural environmental injustices and hydraulic fracturing on Pennsylvania's farms. Journal of Rural Studies 47:278–290; doi:10.1016/j.jrurstud.2015.12.015.

Manson S, Schroeder J, Van Riper D, Ruggles S. 2017. National Historical Geographic Information System: Version 12.0.; doi:10.18128/D050.V12.0.

Mansournia MA, Jewell NP, Greenland S. 2018. Case-control matching: effects, misconceptions, and recommendations. Eur J Epidemiol 33:5–14; doi:10.1007/s10654-017-0325-0.

Maroziene L, Grazuleviciene R. 2002. Maternal exposure to low-level air pollution and pregnancy outcomes: a population-based study. Environmental Health 1:6; doi:10.1186/1476-069X-1-6.

Marrero JE, Townsend-Small A, Lyon DR, Tsai TR, Meinardi S, Blake DR. 2016. Estimating Emissions of Toxic Hydrocarbons from Natural Gas Production Sites in the Barnett Shale Region of Northern Texas. Environ Sci Technol 50:10756–10764; doi:10.1021/acs.est.6b02827.

Martuzzi M, Mitis F, Forastiere F. 2010. Inequalities, inequities, environmental justice in waste management and health. The European Journal of Public Health 20:21–26; doi:10.1093/eurpub/ckp216.

Maskrey JR, Insley AL, Hynds ES, Panko JM. 2016. Air monitoring of volatile organic compounds at relevant receptors during hydraulic fracturing operations in Washington County, Pennsylvania. Environmental Monitoring and Assessment 188; doi:10.1007/s10661-016-5410-4.

Massey DS. 1996. The age of extremes: concentrated affluence and poverty in the twenty-first century. Demography 33: 395–412; discussion 413-416.

McIntosh JC, Ferguson G. 2019. Conventional Oil—The Forgotten Part of the Water-Energy Nexus. Groundwater 57:669–677; doi:10.1111/gwat.12917.

McKenzie LM, Allshouse W, Daniels S. 2019. Congenital heart defects and intensity of oil and gas well site activities in early pregnancy. Environment International 132:104949; doi:10.1016/j.envint.2019.104949.

McKenzie LM, Guo R, Witter RZ, Savitz DA, Newman LS, Adgate JL. 2014. Birth Outcomes and Maternal Residential Proximity to Natural Gas Development in Rural Colorado. Environmental Health Perspectives; doi:10.1289/ehp.1306722.

McKenzie LM, Witter RZ, Newman LS, Adgate JL. 2012a. Human health risk assessment of air emissions from development of unconventional natural gas resources. Science of The Total Environment 424:79–87; doi:10.1016/j.scitotenv.2012.02.018.

McKenzie LM, Witter RZ, Newman LS, Adgate JL. 2012b. Human health risk assessment of air emissions from development of unconventional natural gas resources. Science of The Total Environment 424:79–87; doi:10.1016/j.scitotenv.2012.02.018.

McMahon PB, Barlow JRB, Engle MA, Belitz K, Ging PB, Hunt AG, et al. 2017. Methane and Benzene in Drinking-Water Wells Overlying the Eagle Ford, Fayetteville, and Haynesville Shale Hydrocarbon Production Areas. Environ Sci Technol; doi:10.1021/acs.est.7b00746.

McMahon PB, Vengosh A, Davis TA, Landon MK, Tyne RL, Wright MT, et al. 2019. Occurrence and Sources of Radium in Groundwater Associated with Oil Fields in the Southern San Joaquin Valley, California. Environ Sci Technol 53:9398–9406; doi:10.1021/acs.est.9b02395.

Mellqvist J, Samuelsson J, Andersson P, Brohede S, Isoz O, Ericsson M. 2017. Using Solar Occultation Flux and other Optical Remote Sensing Methodsto measureVOC emissions from a variety of stationary sourcesin the South Coast Air Basin.

Meng Q. 2015a. 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; doi:10.1016/j.scitotenv.2015.02.030.

Meng Q. 2015b. 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; doi:10.1016/j.scitotenv.2015.02.030.

Metzger LF, Landon MK. 2018. Preliminary groundwater salinity mapping near selected oil fields using historical water-sample data, central and southern California. Preliminary groundwater salinity mapping near selected oil fields using historical water-sample data, central and southern California 2018–5082:64; doi:10.3133/sir20185082.

microsoft/USBuildingFootprints. 2019.Microsoft.

Millemann RE, Haynes RJ, Boggs TA, Hildebrand SG, RidgeNationalLaboratory O. 1982. Enhanced oil recovery: environmental issues and state regulatory programs. Environment International 7:165–177; doi:https://doi.org/10.1016/0160-4120(82)90103-9.

Miller A, Siffel C, Correa A. 2010. Residential Mobility During Pregnancy: Patterns and Correlates. Matern Child Health J 14:625–634; doi:10.1007/s10995-009-0492-z.

Miller DJ, Sun K, Tao L, Pan D, Zondlo MA, Nowak JB, et al. 2015. Ammonia and methane dairy emission plumes in the San Joaquin Valley of California from individual feedlot to regional scales. Journal of Geophysical Research: Atmospheres 120:9718–9738; doi:10.1002/2015JD023241.

Mirabelli MC, Wing S, Marshall SW, Wilcosky TC. 2006a. Race, Poverty, and Potential Exposure of Middle-School Students to Air Emissions from Confined Swine Feeding Operations. Environ Health Perspect 114:591–596; doi:10.1289/ehp.8586.

Mirabelli MC, Wing S, Marshall SW, Wilcosky TC. 2006b. Race, Poverty, and Potential Exposure of Middle-School Students to Air Emissions from Confined Swine Feeding Operations. Environ Health Perspect 114:591–596; doi:10.1289/ehp.8586.

Mitka M. 2012. Rigorous Evidence Slim for Determining Health Risks From Natural Gas Fracking. JAMA 307:2135–2136; doi:10.1001/jama.2012.3726.

Mohai P, Bryant B. 1992. Environmental Injustice: Weighing Race and Class as Factors in the Distribution of Environmental Hazards. U Colo L Rev 63: 921.

Mohai P, Pellow D, Roberts JT. 2009. Environmental Justice. Annual Review of Environment and Resources 34:405–430; doi:10.1146/annurev-environ-082508-094348.

Moloney MF, Aycock DM, Cotsonis GA, Myerburg S, Farino C, Lentz M. 2009. An Internet-based migraine headache diary: Issues in Internet-based research. Headache: The Journal of Head and Face Pain 49: 673–686.

Morello-Frosch R, Jesdale BM. 2006. Separate and Unequal: Residential Segregation and Estimated Cancer Risks Associated with Ambient Air Toxics in U.S. Metropolitan Areas. Environ Health Perspect 114:386–393; doi:10.1289/ehp.8500.

Morello-Frosch R, Jesdale BM, Sadd JL, Pastor M. 2010. Ambient air pollution exposure and full-term birth weight in California. Environmental Health 9:44; doi:10.1186/1476-069X-9-44.

Morello-Frosch RA. 2002. Discrimination and the Political Economy of Environmental Inequality. Environment and Planning C: Government and Policy 20:477–496; doi:10.1068/c03r.

Moster D, Lie RT, Markestad T. 2008. Long-Term Medical and Social Consequences of Preterm Birth. New England Journal of Medicine 359:262–273; doi:10.1056/NEJMoa0706475.

Mukamal KJ, Wellenius GA, Suh HH, Mittleman MA. 2009. Weather and air pollution as triggers of severe headaches. Neurology 72: 922–927.

Mukerjee S, Smith LA, Thoma ED, Whitaker DA, Oliver KD, Duvall R, et al. 2020. Spatial analysis of volatile organic compounds using passive samplers in the Rubbertown industrial area of Louisville, Kentucky, USA. Atmospheric Pollution Research 11:81–86; doi:10.1016/j.apr.2020.02.021.

Myhre G, Shindell D, Bréon F-M, Collins W, Fuglestvedt J, Huang J, et al. 2013. 2013: Anthropogenic and Natural Radiative Forcing. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, et al., eds). Cambridge University Press:Cambridge, United Kingdom and New York, NY, USA.

Nationwide Parcel Data & Property Level Geocodes | SmartParcels®. Digital Map Products. Available: https://www.digmap.com/platform/smartparcels/ [accessed 1 July 2020].

Neuman JA, Nowak JB, Brock CA, Trainer M, Fehsenfeld FC, Holloway JS, et al. 2003. Variability in ammonium nitrate formation and nitric acid depletion with altitude and location over California. Journal of Geophysical Research: Atmospheres 108; doi:10.1029/2003JD003616.

Ngwabie NM, Jeppsson K-H, Nimmermark S, Swensson C, Gustafsson G. 2009. Multi-location measurements of greenhouse gases and emission rates of methane and ammonia from a naturally-ventilated barn for dairy cows. Biosystems Engineering 103:68–77; doi:10.1016/j.biosystemseng.2009.02.004.

OEHHA (CA Office of Environmental Health Hazard Assessment). 2016. CalEnviroScreen 3.0. OEHHA. Available: https://oehha.ca.gov/calenviroscreen/report/calenviroscreen-30 [accessed 23 March 2020].

Ogneva-Himmelberger Y, Huang L. 2015a. Spatial distribution of unconventional gas wells and human populations in the Marcellus Shale in the United States: Vulnerability analysis. Applied Geography 60:165–174; doi:10.1016/j.apgeog.2015.03.011.

Ogneva-Himmelberger Y, Huang L. 2015b. Spatial distribution of unconventional gas wells and human populations in the Marcellus Shale in the United States: Vulnerability analysis. Applied Geography 60:165–174; doi:10.1016/j.apgeog.2015.03.011.

O'Neill MS, Jerrett M, Kawachi K, Levy JI, J AJ, Nelson G, et al. 2003. Health, wealth, and air pollution: advancing theory and methods. Environmental Health Perspectives 111:1861–1870; doi:10.1289/ehp.6334.

Owusu JT, Anderson FJ, Coleman J, Oppong S, Seffah JD, Aikins A, et al. 2013. Association of maternal sleep practices with pre-eclampsia, low birth weight, and stillbirth among Ghanaian women. International Journal of Gynecology & Obstetrics 121:261–265; doi:10.1016/j.ijgo.2013.01.013.

Pace C, Balazs C, Cushing L, Morello-Frosch R. 2020a. UC Berkeley Water Equity Science Shop Community Water System boundaries version 1.0, 2019.

Pace C, Balazs C, Cushing LJ, Morello-Frosch R. 2020b. UC Berkeley Water Equity Science Shop domestic well communities version 1.0, 2019.

Padilla CM, Kihal-Talantikite W, Vieira VM, Rossello P, Nir GL, Zmirou-Navier D, et al. 2014. Air quality and social deprivation in four French metropolitan areas—A localized spatio-temporal environmental inequality analysis. Environmental Research 134:315–324; doi:10.1016/j.envres.2014.07.017.

Padula AM, Mortimer KM, Tager IB, Hammond SK, Lurmann FW, Yang W, et al. 2014. Traffic-Related Air Pollution and Risk of Preterm Birth in the San Joaquin Valley of California. Ann Epidemiol 24:888-95e4; doi:10.1016/j.annepidem.2014.10.004.

Palagini L, Gemignani A, Banti S, Manconi M, Mauri M, Riemann D. 2014. Chronic sleep loss during pregnancy as a determinant of stress: impact on pregnancy outcome. Sleep Medicine 15:853–859; doi:10.1016/j.sleep.2014.02.013.

Pavley F. 2013. Bill Text - SB-4 Oil and gas: well stimulation. Available: https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201320140SB4 [accessed 7 June 2020].

Peatfield R, Glover V, Littlewood J, Sandler M, Clifford Rose F. 1984. The prevalence of diet-induced migraine. Cephalalgia 4: 179–183.

Pennington AF, Strickland MJ, Klein M, Zhai X, Russell AG, Hansen C, et al. 2017. Measurement error in mobile source air pollution exposure estimates due to residential mobility during pregnancy. Journal of Exposure Science and Environmental Epidemiology 27:513–520; doi:10.1038/jes.2016.66.

Ponce NA, Hoggatt KJ, Wilhelm M, Ritz B. 2005. Preterm Birth: The Interaction of Traffic-related Air Pollution with Economic Hardship in Los Angeles Neighborhoods. Am J Epidemiol 162:140–148; doi:10.1093/aje/kwi173.

Pope CA 3rd, Hansen ML, Long RW, Nielsen KR, Eatough NL, Wilson WE, et al. 2004. Ambient particulate air pollution, heart rate variability, and blood markers of inflammation in a panel of elderly subjects. Environ Health Perspect 112:339–45; doi:10.1289/ehp.6588.

Pope CA, Burnett RT, Krewski D, Jerrett M, Shi Y, Calle EE, et al. 2009. Cardiovascular Mortality and Exposure to Airborne Fine Particulate Matter and Cigarette Smoke: Shape of the Exposure-Response Relationship. Circulation 120:941–948; doi:10.1161/CIRCULATIONAHA.109.857888.

Porter JK, Di Tanna GL, Lipton RB, Sapra S, Villa G. 2019. Costs of acute headache medication use and productivity losses among patients with migraine: insights from three randomized controlled trials. PharmacoEconomics-open 3: 411–417.

Pressman A, Jacobson A, Eguilos R, Gelfand A, Huynh C, Hamilton L, et al. 2016. Prevalence of migraine in a diverse community—electronic methods for migraine ascertainment in a large integrated health plan. Cephalalgia 36: 325–334.

Prince PB, Rapoport AM, Sheftell FD, Tepper SJ, Bigal ME. 2004. The effect of weather on headache. Headache: The Journal of Head and Face Pain 44: 596–602.

Rasmussen SG, Ogburn EL, McCormack M, Casey JA, Bandeen-Roche K, Mercer DG, et al. 2016. Association between unconventional natural gas development in the marcellus shale and asthma exacerbations. JAMA Intern Med 176:1334–43; doi:10.1001/jamainternmed.2016.2436.

Raz R, Kioumourtzoglou M-A, Weisskopf MG. 2018. Live-Birth Bias and Observed Associations Between Air Pollution and Autism. American Journal of Epidemiology 187:2292–2296; doi:10.1093/aje/kwy172.

Rich AL, Orimoloye HT. 2016. Elevated Atmospheric Levels of Benzene and Benzene-Related Compounds from Unconventional Shale Extraction and Processing: Human Health Concern for Residential Communities. Environ Health Insights 10:75–82; doi:10.4137/EHI.S33314.

Ricketts TC. 2000. The Changing Nature of Rural Health Care. Annual Review of Public Health 21:639–657; doi:10.1146/annurev.publhealth.21.1.639.

Ringquist EJ. 2005. Assessing evidence of environmental inequities: A meta-analysis. Journal of Policy Analysis and Management 24:223–247; doi:10.1002/pam.20088.

Ritz B, Wilhelm M. 2008. Ambient Air Pollution and Adverse Birth Outcomes: Methodologic Issues in an Emerging Field. Basic & Clinical Pharmacology & Toxicology 102:182–190; doi:10.1111/j.1742-7843.2007.00161.x.

Ritz B, Wilhelm M, Hoggatt KJ, Ghosh JKC. 2007. Ambient Air Pollution and Preterm Birth in the Environment and Pregnancy Outcomes Study at the University of California, Los Angeles. Am J Epidemiol 166:1045–1052; doi:10.1093/aje/kwm181.

Rondó PHC, Ferreira RF, Nogueira F, Ribeiro MCN, Lobert H, Artes R. 2003. Maternal psychological stress and distress as predictors of low birth weight, prematurity and intrauterine growth retardation. Eur J Clin Nutr 57:266–272; doi:10.1038/sj.ejcn.1601526.

Rothstein R. 2017. *The color of law: a forgotten history of how our government segregated America*. First edition. Liveright Publishing Corporation, a division of W. W. Norton & Company:New York ; London.

Roy AA, Adams PJ, Robinson AL. 2014a. Air pollutant emissions from the development, production, and processing of Marcellus Shale natural gas. Journal of the Air & Waste Management Association 64:19–37; doi:10.1080/10962247.2013.826151.

Roy AA, Adams PJ, Robinson AL. 2014b. Air pollutant emissions from the development, production, and processing of Marcellus Shale natural gas. Journal of the Air & Waste Management Association 64:19–37; doi:10.1080/10962247.2013.826151.

Roy AA, Adams PJ, Robinson AL. 2014c. Air pollutant emissions from the development, production, and processing of Marcellus Shale natural gas. Journal of the Air & Waste Management Association 64:19–37; doi:10.1080/10962247.2013.826151.

Rusconi F, Catelan D, Accetta G, Peluso M, Pistelli R, Barbone F, et al. 2011. Asthma Symptoms, Lung Function, and Markers of Oxidative Stress and Inflammation in Children Exposed to Oil Refinery Pollution. Journal of Asthma 48:84–90; doi:10.3109/02770903.2010.538106.

Saha R, Mohai P. 2005. Historical Context and Hazardous Waste Facility Siting: Understanding Temporal Patterns in Michigan. Social Problems 52:618–648; doi:10.1525/sp.2005.52.4.618.

Saigal S, Doyle LW. 2008. An overview of mortality and sequelae of preterm birth from infancy to adulthood. 371: 9.

San Sebastian M, Armstrong B, Stephens C. 2002. Outcomes of Pregnancy among Women Living in the Proximity of Oil Fields in the Amazon Basin of Ecuador. INT J OCCUP ENVIRON HEALTH 8: 312–319.

Sanchez NP, Saffari A, Barczyk S, Coleman BK, Naufal Z, Rabideau C, et al. 2019. Results of Three Years of Ambient Air Monitoring Near a Petroleum Refinery in Richmond, California, USA. Atmosphere 10:385; doi:10.3390/atmos10070385.

SAS. PROC VARIOGRAM: Aspects of Semivariogram Model Fitting. SAS/STAT(R) 922 User’s Guide. Available: https://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#statug_variogram_a0000000605.htm [accessed 26 October 2018].

SCAQMD (South Coast Air Quality Management District). 2019. Community Emissions Reduction Plan: Chapter 5 Oil Drilling and Production, Wilmington, Carson, West Long Beach.

Schatzki T, Stavins RN. 2009. Addressing Environmental Justice Concerns in the Design of California’s Climate Policy. 33.

Schmoker JW, Klett TR. 2005. Petroleum Systems and Geologic Assessment of Oil and Gas in the Southwestern Wyoming Province, Wyoming, Colorado, and Utah: Chapter 19, U.S. Geological Survey Assessment Concepts for Conventional Petroleum Accumulations.

Schraufnagel DE. 2020. The health effects of ultrafine particles. Experimental & Molecular Medicine 1–7.

Schreiber-Gregory DN. 2012. Paper 1404-2017 - Multicollinearity: What Is It, Why Should We Care, and How Can It Be Controlled?

Shonkoff S, Morello-Frosch R, Pastor M, Sadd J. 2011. Environmental Health and Equity Implications of Climate Change and Mitigation Policies in California: A Review of the Literature. Climatic Change 109: S485-503.

Shonkoff SBC, DiGiulio DC. 2019. Reasons to maintain a definition of protected groundwater equivalent to an USDW during well stimulation and oil and gas development in California.

Shonkoff SBC, Domen JK, Hill LAL. 2019. Human health and oil and gas development: An assessment of chemical usage in oil and gas activities in the Los Angeles Basin and the City of Los Angeles.

Silva GS, Warren JL, Deziel NC. 2018. Spatial Modeling to Identify Sociodemographic Predictors of Hydraulic Fracturing Wastewater Injection Wells in Ohio Census Block Groups. Environmental Health Perspectives 126; doi:10.1289/EHP2663.

Singh GK, Siahpush M. 2014. Widening Rural–Urban Disparities in Life Expectancy, U.S., 1969–2009. American Journal of Preventive Medicine 46:e19–e29; doi:10.1016/j.amepre.2013.10.017.

Smargiassi Audrey, Kosatsky Tom, Hicks John, Plante Céline, Armstrong Ben, Villeneuve Paul J., et al. 2009. Risk of Asthmatic Episodes in Children Exposed to Sulfur Dioxide Stack Emissions from a Refinery Point Source in Montreal, Canada. Environmental Health Perspectives 117:653–659; doi:10.1289/ehp.0800010.

Smith KR, Peel JL. 2010. Mind the Gap. Environmental Health Perspectives 118:1643–1645; doi:10.1289/ehp.1002517.

Smith R, Knight R, Fendorf S. 2018. Overpumping leads to California groundwater arsenic threat. Nature Communications 9; doi:10.1038/s41467-018-04475-3.

Soelter AD, Miller EG. 1999. Capacity development: the small system perspective. Journal - AWWA 91:110–122; doi:10.1002/j.1551-8833.1999.tb08617.x.

Spielman SE, Yoo E-H, Linkletter C. 2013. Neighborhood contexts, health, and behavior: understanding the role of scale and residential sorting. Environment and Planning B: Planning and Design 40: 489–506.

Spierings EL, Ranke AH, Honkoop PC. 2001. Precipitating and aggravating factors of migraine versus tension-type headache. Headache: The journal of head and face pain 41: 554–558.

Srebotnjak, Tanja, Rotkin-Ellman M. 2014. Drilling in California: Who's at risk? 28.

Stacy SL, Brink LL, Larkin JC, Sadovsky Y, Goldstein BD, Pitt BR, et al. 2015. Perinatal Outcomes and Unconventional Natural Gas Operations in Southwest Pennsylvania. J. Meliker, ed PLOS ONE 10:e0126425; doi:10.1371/journal.pone.0126425.

Staines A Crowley, D, Bruen, M and O'Connor, P. 2004. Public Health and Landfill Sites.

Statewide Database | Election Data. Available: https://statewidedatabase.org/election.html [accessed 1 July 2020].

Stieb DM, Chen L, Eshoul M, Judek S. 2012. Ambient air pollution, birth weight and preterm birth: A systematic review and meta-analysis. Environmental Research 117:100–111; doi:10.1016/j.envres.2012.05.007.

Stovner LJ, Nichols E, Steiner TJ, Abd-Allah F, Abdelalim A, Al-Raddadi RM, et al. 2018. Global, regional, and national burden of migraine and tension-type headache, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet Neurology 17: 954–976.

Su JG, Larson T, Gould T, Cohen M, Buzzelli M. 2010. Transboundary air pollution and environmental justice: Vancouver and Seattle compared. GeoJournal 75:595–608; doi:10.1007/s10708-009-9269-6.

Sun P, Young B, Elgowainy A, Lu Z, Wang M, Morelli B, et al. 2019. Criteria Air Pollutant and Greenhouse Gases Emissions from U.S. Refineries Allocated to Refinery Products. Environ Sci Technol 53:6556–6569; doi:10.1021/acs.est.8b05870.

Swarthout RF, Russo RS, Zhou Y, Hart AH, Sive BC. 2013. Volatile organic compound distributions during the NACHTT campaign at the Boulder Atmospheric Observatory: Influence of urban and natural gas sources. J Geophys Res Atmos 118:10,614-10,637; doi:10.1002/jgrd.50722.

SWRCB (CA State Water Resources Control Board). 2015a. A Guide for Private Domestic Well Owners.

SWRCB (CA State Water Resources Control Board). 2019. GAMA Groundwater. Available: https://gamagroundwater.waterboards.ca.gov/gama/gamamap/public/ [accessed 13 November 2019].

SWRCB (CA State Water Resources Control Board). 2012. Public Water System Definitions.

SWRCB (CA State Water Resources Control Board). 2015b. Water Quality in Areas of Oil and Gas Production – Regional Groundwater Monitoring. Available: https://www.waterboards.ca.gov/water_issues/programs/groundwater/sb4/regional_monitoring/ [accessed 15 November 2019].

Szyszkowicz M, Kaplan G, Grafstein E, Rowe B. 2009a. Emergency department visits for migraine and headache: a multi-city study. International Journal of Occupational Medicine and Environmental Health 22: 235–242.

Szyszkowicz M, Rowe B, Kaplan G. 2009b. Ambient sulphur dioxide exposure and emergency department visits for migraine in Vancouver, Canada. International Journal of Occupational Medicine and Environmental Health 22: 7–12.

Szyszkowicz M, Stieb DM, Rowe BH. 2009c. Air pollution and daily ED visits for migraine and headache in Edmonton, Canada. The American journal of emergency medicine 27: 391–396.

Talge NM, Mudd LM, Sikorskii A, Basso O. 2014a. United States Birth Weight Reference Corrected For Implausible Gestational Age Estimates. Pediatrics 133:844–853; doi:10.1542/peds.2013-3285.

Talge NM, Mudd LM, Sikorskii A, Basso O. 2014b. United States Birth Weight Reference Corrected For Implausible Gestational Age Estimates. Pediatrics 133:844–853; doi:10.1542/peds.2013-3285.

Teitler JO, Reichman NE, Nepomnyaschy L, Martinson M. 2007. A Cross-National Comparison of Racial and Ethnic Disparities in Low Birth Weight in the United States and England. PEDIATRICS 120:e1182–e1189; doi:10.1542/peds.2006-3526.

Thompson CR, Hueber J, Helmig D. 2014. Influence of oil and gas emissions on ambient atmospheric non-methane hydrocarbons in residential areas of Northeastern Colorado. Elementa: Science of the Anthropocene 2; doi:10.12952/journal.elementa.000035.

Thurmond TK. 2016. California Assembly Bill 1496: Methane emissions.

Tran KV, Casey JA, Cushing LJ, Morello-Frosch R. 2020a. Residential Proximity to Oil and Gas Development and Birth Outcomes in California: A Retrospective Cohort Study of 2006–2015 Births. Environmental Health Perspectives 128:0670011–06700113; doi:10.1289/EHP5842.

Tran KV, Casey JA, Cushing LJ, Morello-Frosch RA. 2020b. Residential Proximity to Oil and Gas Development and Birth Outcomes in California: A Retrospective Cohort Study of 2006–2015 Births. Environmental Health Perspectives 128:067001–0670013; doi:10.1289/EHP5842.

Turner L, Molgaard C, Gardner C, Rothrock J, Stang P. 1995. Migraine trigger factors in a non-clinical Mexican-American population in San Diego county: implications for etiology. Cephalalgia 15: 523–530.

Tustin AW, Hirsch AG, Rasmussen SG, Casey JA, Bandeen-Roche K, Schwartz BS. 2017a. Associations between Unconventional Natural Gas Development and Nasal and Sinus, Migraine Headache, and Fatigue Symptoms in Pennsylvania. Environ Health Perspect 125:189–197; doi:10.1289/EHP281.

Tustin AW, Hirsch AG, Rasmussen SG, Casey JA, Bandeen-Roche K, Schwartz BS. 2017b. Associations between unconventional natural gas development and nasal and sinus, migraine headache, and fatigue symptoms in Pennsylvania. Environmental health perspectives 125: 189–197.

UCLA: Statistical Consulting Group. Analyzing and Visualizing Interactions in SAS. Available: https://stats.idre.ucla.edu/sas/seminars/analyzing-and-visualizing-interactions/ [accessed 20 October 2019].

US BLM (Bureau of Land Management). 2006. Final environmental impact statement: Jonah infill drilling project, Sublette County, Wyoming.

US Census Bureau. 2010 Census Urban and Rural Classification and Urban Area Criteria. Available: https://www.census.gov/geo/reference/ua/urban-rural-2010.html [accessed 1 March 2019a].

US Census Bureau. Urban and Rural classification. The United States Census Bureau. Available: https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural.html [accessed 24 January 2021b].

US EIA (Energy Information Agency). 2018a. CA - State Profile and Energy Estimates. Available: https://www.eia.gov/state/analysis.cfm?sid=CA [accessed 6 November 2016].

US EIA (Energy Information Agency). 2019. Layer Information for Interactive State Maps. Available: https://www.eia.gov/maps/layer_info-m.php [accessed 7 May 2020].

US EIA (Energy Information Agency). 2018b. U.S. Crude Oil, Natural Gas, and Natural Gas Liquids Proved Reserves. Available: http://www.eia.gov/naturalgas/crudeoilreserves/ [accessed 6 November 2016].

US EIA (Energy Information Agency). 2020. U.S. States - Rankings. Available: https://www.eia.gov/state/rankings/#/series/46 [accessed 16 June 2020].

U.S. Environmental Protection Agency. 2020. Integrated Science Assessment (ISA) for Ozone and Related Photochemical Oxidants (Final Report). 1468.

US EPA (Environmental Protection Agency). 2015a. Analysis of Hydraulic Fracturing Fluid Data from the FracFocus Chemical Disclosure Registry 1.0. 168.

US EPA (Environmental Protection Agency). 2017. Aquifer Exemption Data. US EPA. Available: https://www.epa.gov/uic/aquifer-exemption-data [accessed 13 November 2019].

US EPA (Environmental Protection Agency). 2015b. Class II Oil and Gas Related Injection Wells. US EPA. Available: https://www.epa.gov/uic/class-ii-oil-and-gas-related-injection-wells [accessed 6 June 2020].

US EPA (Environmental Protection Agency). 2015c. EPA Oversight of California's Underground Injection Control (UIC) Program. US EPA. Available: https://www.epa.gov/uic/epa-oversight-californias-underground-injection-control-uic-program [accessed 11 November 2019].

US EPA (Environmental Protection Agency). 2016. Hydraulic Fracturing for Oil and Gas: Impacts from the Hydraulic Fracturing Water Cycle on Drinking Water Resources in the United States (Final Report). US Environmental Protection Agency, Washington, DC, EPA/600/R-16/236F 666.

US EPA (Environmental Protection Agency). 2018. Inventory of U.S. Greenhouse Gas Emissions and Sinks 1990-2016: Abandoned Oil and Gas Wells.

Valero de Bernabé J, Soriano T, Albaladejo R, Juarranz M, Calle ME, Martínez D, et al. 2004. Risk factors for low birth weight: a review. European Journal of Obstetrics & Gynecology and Reproductive Biology 116:3–15; doi:10.1016/j.ejogrb.2004.03.007.

van Kesteren P, in Biesebeek J, Palmen N, Bakker J, Muller J. 2013. Risk assessment of an increased concentration limit of benzene in natural gas. 36.

Veil JA, Puder MG, Elcock D, Redweik, R. J., Jr. 2004. A white paper describing produced water from production of crude oil, natural gas, and coal bed methane.; doi:10.2172/821666.

Vengosh A, Jackson RB, Warner N, Darrah TH, Kondash A. 2014a. A Critical Review of the Risks to Water Resources from Unconventional Shale Gas Development and Hydraulic Fracturing in the United States. Environmental Science & Technology 48:8334–8348; doi:10.1021/es405118y.

Vengosh A, Jackson RB, Warner N, Darrah TH, Kondash A. 2014b. A Critical Review of the Risks to Water Resources from Unconventional Shale Gas Development and Hydraulic Fracturing in the United States. Environ Sci Technol 48:8334–8348; doi:10.1021/es405118y.

Viehe FW. 1981. Black Gold Suburbs: The Influence of the Extractive Industry on the Suburbanization of Los Angeles, 1890-1930. Journal of Urban History 8:3–26; doi:10.1177/009614428100800101.

von Mackensen S, Hoeppe P, Maarouf A, Tourigny P, Nowak D. 2005. Prevalence of weather sensitivity in Germany and Canada. International journal of biometeorology 49: 156–166.

Vrijheid M. 2000. Health effects of residence near hazardous waste landfill sites: a review of epidemiologic literature. Environmental health perspectives 108 Suppl 1: 101–12.

Walker JD. 2011. California Class II Underground Injection Control Program Review.

Walker Whitworth K, Kaye Marshall A, Symanski E. 2018. Drilling and Production Activity Related to Unconventional Gas Development and Severity of Preterm Birth. Environmental Health Perspectives 126; doi:10.1289/EHP2622.

Warneke C, Geiger F, Edwards PM, Dube W, Pétron G, Kofler J, et al. 2014. Volatile organic compound emissions from the oil and natural gas industry in the Uintah Basin, Utah: oil and gas well pad emissions compared to ambient air composition. Atmospheric Chemistry and Physics 14:10977–10988; doi:10.5194/acp-14-10977-2014.

Watson T. 2009. Inequality and the measurement of residential segregation by income in American neighborhoods. Review of Income and Wealth 55: 820–844.

West JJ, Fiore AM, Horowitz LW, Mauzerall DL. 2006. Global health benefits of mitigating ozone pollution with methane emission controls. PNAS 103:3988–3993; doi:10.1073/pnas.0600201103.

White N, teWaterNaude J, van der Walt A, Ravenscroft G, Roberts W, Ehrlich R. 2009. Meteorologically estimated exposure but not distance predicts asthma symptoms in schoolchildren in the environs of a petrochemical refinery: a cross-sectional study. Environ Health 8:45; doi:10.1186/1476-069X-8-45.

Whitworth KW, Marshall AK, Symanski E. 2017. Maternal residential proximity to unconventional gas development and perinatal outcomes among a diverse urban population in Texas. PLOS ONE 12:e0180966; doi:10.1371/journal.pone.0180966.

Wilson SM, Fraser-Rahim H, Williams E, Zhang H, Rice L, Svendsen E, et al. 2012. Assessment of the Distribution of Toxic Release Inventory Facilities in Metropolitan Charleston: An Environmental Justice Case Study. Am J Public Health 102:1974–1980; doi:10.2105/AJPH.2012.300700.

Wing S, Cole D, Grant G. 2000a. Environmental injustice in North Carolina's hog industry. Environ Health Perspect 108: 225–231.

Wing S, Cole D, Grant G. 2000b. Environmental injustice in North Carolina's hog industry. Environ Health Perspect 108: 225–231.

Win-Shwe T-T, Fujimaki H. 2011. Nanoparticles and neurotoxicity. International journal of molecular sciences 12: 6267–6280.

Wöber C, Holzhammer J, Zeitlhofer J, Wessely P, Wöber-Bingöl Ç. 2006. Trigger factors of migraine and tension-type headache: experience and knowledge of the patients. The journal of headache and pain 7: 188–195.

Woo B, Kravitz-Wirtz N, Sass V, Crowder K, Teixeira S, Takeuchi DT. 2019. Residential segregation and racial/ethnic disparities in ambient air pollution. Race and social problems 11: 60–67.

Woodruff TJ, Parker JD, Darrow LA, Slama R, Bell ML, Choi H, et al. 2009. Methodological issues in studies of air pollution and reproductive health. Environmental Research 109:311–320; doi:10.1016/j.envres.2008.12.012.

Wright MT, McMahon PB, Landon MK, Kulongoski JT. 2019. Groundwater quality of a public supply aquifer in proximity to oil development, Fruitvale oil field, Bakersfield, California. Applied Geochemistry 106:82–95; doi:10.1016/j.apgeochem.2019.05.003.

Wu W, Zhang G, Kai P. 2012. Ammonia and methane emissions from two naturally ventilated dairy cattle buildings and the influence of climatic factors on ammonia emissions. Atmospheric Environment 61:232–243; doi:10.1016/j.atmosenv.2012.07.050.

Wunderli S, Gehrig R. 1990. Surface ozone in rural, urban and alpine regions of switzerland. Atmospheric Environment Part A General Topics 24:2641–2646; doi:10.1016/0960-1686(90)90143-B.

Yang AC, Fuh J-L, Huang NE, Shia B-C, Peng C-K, Wang S-J. 2011. Temporal associations between weather and headache: analysis by empirical mode decomposition. PloS one 6: e14612.

Yang CY, Cheng BH, Hsu TY, Tsai SS, Hung CF, Wu TN. 2000. Female lung cancer mortality and sex ratios at birth near a petroleum refinery plant. Environ Res 83:33–40; doi:10.1006/enrs.2000.4038.

Yu C-L, Wang S-F, Pan P-C, Wu M-T, Ho C-K, Smith TJ, et al. 2006. Residential exposure to petrochemicals and the risk of leukemia: using geographic information system tools to estimate individual-level residential exposure. Am J Epidemiol 164:200–207; doi:10.1093/aje/kwj182.

Zavala-Araiza D, Alvarez RA, Lyon DR, Allen DT, Marchese AJ, Zimmerle DJ, et al. 2017. Super-emitters in natural gas infrastructure are caused by abnormal process conditions. Nat Commun 8; doi:10.1038/ncomms14012.

Zavala-Araiza D, Lyon D, Alvarez RA, Palacios V, Harriss R, Lan X, et al. 2015. Toward a Functional Definition of Methane Super-Emitters: Application to Natural Gas Production Sites. Environ Sci Technol 49:8167–8174; doi:10.1021/acs.est.5b00133.

Zhao X, Zhang X, Xu X, Xu J, Meng W, Pu W. 2009. Seasonal and diurnal variations of ambient PM2.5 concentration in urban and rural environments in Beijing. Atmospheric Environment 43:2893–2900; doi:10.1016/j.atmosenv.2009.03.009.

Zhu X, Liu Y, Chen Y, Yao C, Che Z, Cao J. 2015. Maternal exposure to fine particulate matter (PM2.5) and pregnancy outcomes: a meta-analysis. Environ Sci Pollut Res 22:3383–3396; doi:10.1007/s11356-014-3458-7.

Zwickl K. 2019. The demographics of fracking: A spatial analysis for four U.S. states. Ecological Economics 161:202–215; doi:10.1016/j.ecolecon.2019.02.001.