Impact of proximity to gas production activity on birth outcomes across the US, 2005–2018

Christopher F Baum (Boston College, DIW Berlin & CESIS)
Hailee Schuele (Boston College)
Philip J. Landrigan (Boston College)
Summer Sherburne Hawkins (Boston College)

London Stata Conference, September 2022

Supported by a Public Policy Award from the American Lung Association to S.S. Hawkins

Introduction

We evaluate the associations between US natural gas production activity, in recent years often involving hydraulic fracturing, or 'fracking', and birth outcomes for mothers with prenatal exposure to production facilities.

Although previous research focused on high-producing states has analyzed these health effects, our study is more comprehensive, including detailed production levels of over 1 million natural gas development facilities in 28 US states, with their locations geocoded within 1,984 US counties.

The monthly production data, distance-weighted and aggregated to the county level, are linked to 9 prenatal months for over 33 million singleton births in the 2005-2018 period for mothers living in US counties.

Introduction

We evaluate the associations between US natural gas production activity, in recent years often involving hydraulic fracturing, or 'fracking', and birth outcomes for mothers with prenatal exposure to production facilities.

Although previous research focused on high-producing states has analyzed these health effects, our study is more comprehensive, including detailed production levels of over 1 million natural gas development facilities in 28 US states, with their locations geocoded within 1,984 US counties.

The monthly production data, distance-weighted and aggregated to the county level, are linked to 9 prenatal months for over 33 million singleton births in the 2005-2018 period for mothers living in US counties.

Introduction

We evaluate the associations between US natural gas production activity, in recent years often involving hydraulic fracturing, or 'fracking', and birth outcomes for mothers with prenatal exposure to production facilities.

Although previous research focused on high-producing states has analyzed these health effects, our study is more comprehensive, including detailed production levels of over 1 million natural gas development facilities in 28 US states, with their locations geocoded within 1,984 US counties.

The monthly production data, distance-weighted and aggregated to the county level, are linked to 9 prenatal months for over 33 million singleton births in the 2005-2018 period for mothers living in US counties.

Context for US natural gas production

While US conventional natural gas production declined from 2005–2019, that produced by hydraulic fracturing, or 'fracking', increased by 275% over that period, now representing 75% of the domestic natural gas supply.

In 2017, it was estimated that 17.6 million people in the US lived within one mile of an active gas and/or oil well. While this boom in gas and oil production can lead to short-term economic benefits such as domestic growth and affordable fuel, it reduces the incentives to deal with climate change. In addition, there is mounting evidence of negative effects on public health.

To illustrate the trends in NGD production and its geographic spread in the continental US:

Context for US natural gas production

While US conventional natural gas production declined from 2005–2019, that produced by hydraulic fracturing, or 'fracking', increased by 275% over that period, now representing 75% of the domestic natural gas supply.

In 2017, it was estimated that 17.6 million people in the US lived within one mile of an active gas and/or oil well. While this boom in gas and oil production can lead to short-term economic benefits such as domestic growth and affordable fuel, it reduces the incentives to deal with climate change. In addition, there is mounting evidence of negative effects on public health.

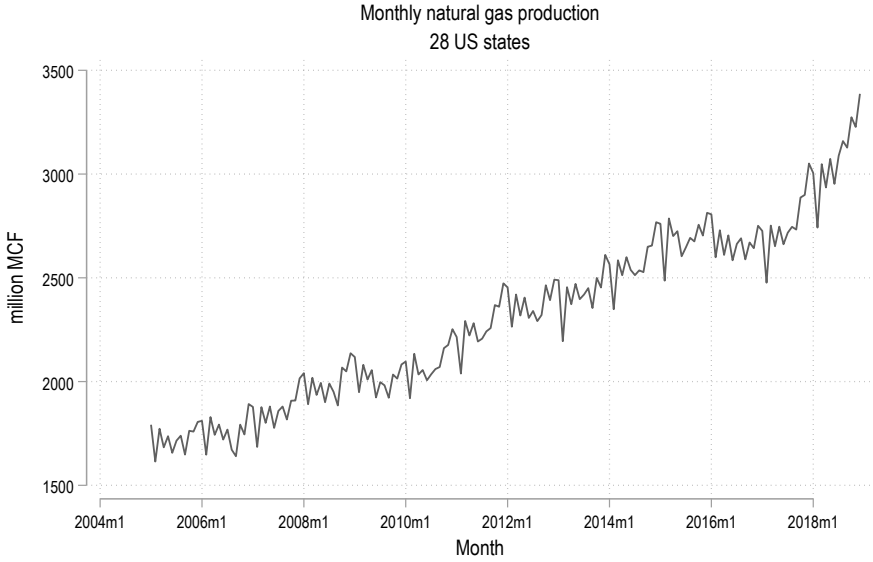
To illustrate the trends in NGD production and its geographic spread in the continental US:

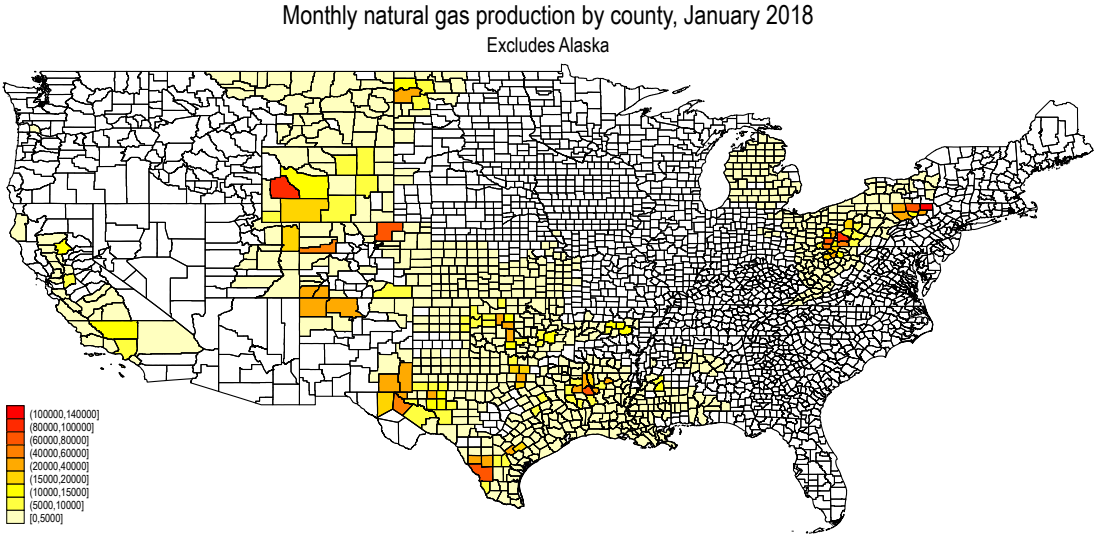
Context for US natural gas production

While US conventional natural gas production declined from 2005–2019, that produced by hydraulic fracturing, or 'fracking', increased by 275% over that period, now representing 75% of the domestic natural gas supply.

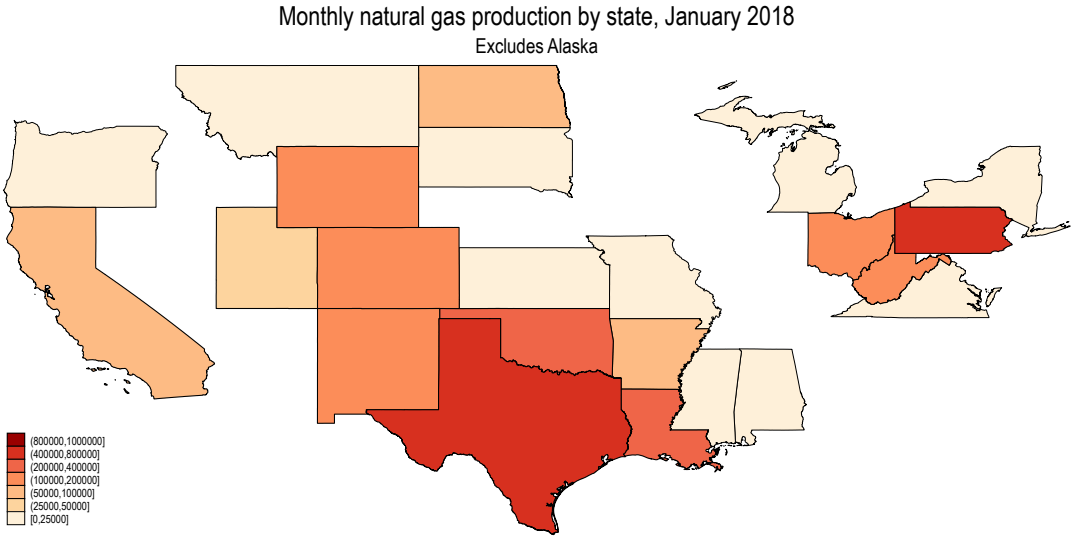
In 2017, it was estimated that 17.6 million people in the US lived within one mile of an active gas and/or oil well. While this boom in gas and oil production can lead to short-term economic benefits such as domestic growth and affordable fuel, it reduces the incentives to deal with climate change. In addition, there is mounting evidence of negative effects on public health.

To illustrate the trends in NGD production and its geographic spread in the continental US:


Monthly natural gas production
28 US states
million MCF
3500
3000
2500
2000
1500
2004m1
2006m1
2008m1
2010m1
2012m1
2014m1
2016m1
2018m1
Month


Monthly natural gas production by county, January 2018
Excludes Alaska
(100000,140000]
(80000,100000]
(60000,80000]
(40000,60000]
(20000,40000]
(15000,20000]
(10000,15000]
(5000,10000]
[0,5000]

From the perspective of states with NGD production:



Prior public health studies

Research over the last decade has demonstrated adverse effects on infant health of proximity to natural gas development (NGD) sites, gauged by distance of residence from gas wells and/or volume of well production.

A review in 2020 evaluated 12 studies conducted in single localities examining the impact of in utero exposure to NGD on infant health.

Four studies of seven focused on fetal growth found evidence of lower birth weight and a higher likelihood to be born small for gestational age (SGA). Five studies found a significant increase in the risk of preterm delivery. More recent studies have found similar associations with being born low birth weight (LBW), SGA or preterm.

Prior public health studies

Research over the last decade has demonstrated adverse effects on infant health of proximity to natural gas development (NGD) sites, gauged by distance of residence from gas wells and/or volume of well production.

A review in 2020 evaluated 12 studies conducted in single localities examining the impact of in utero exposure to NGD on infant health.

Four studies of seven focused on fetal growth found evidence of lower birth weight and a higher likelihood to be born small for gestational age (SGA). Five studies found a significant increase in the risk of preterm delivery. More recent studies have found similar associations with being born low birth weight (LBW), SGA or preterm.

Prior public health studies

Research over the last decade has demonstrated adverse effects on infant health of proximity to natural gas development (NGD) sites, gauged by distance of residence from gas wells and/or volume of well production.

A review in 2020 evaluated 12 studies conducted in single localities examining the impact of in utero exposure to NGD on infant health.

Four studies of seven focused on fetal growth found evidence of lower birth weight and a higher likelihood to be born small for gestational age (SGA). Five studies found a significant increase in the risk of preterm delivery. More recent studies have found similar associations with being born low birth weight (LBW), SGA or preterm.

Those at greater risk to environmental burdens are often living in disadvantaged or minoritized communities. Some studies have found that infants born to Hispanic or Black women, women of lower socioeconomic circumstances, or women with lower educational attainmment had greater reductions in birth weight, and increased risk of LBW, SGA or preterm birth with NGD exposure than women in more favorable socioeconomic circumstances.

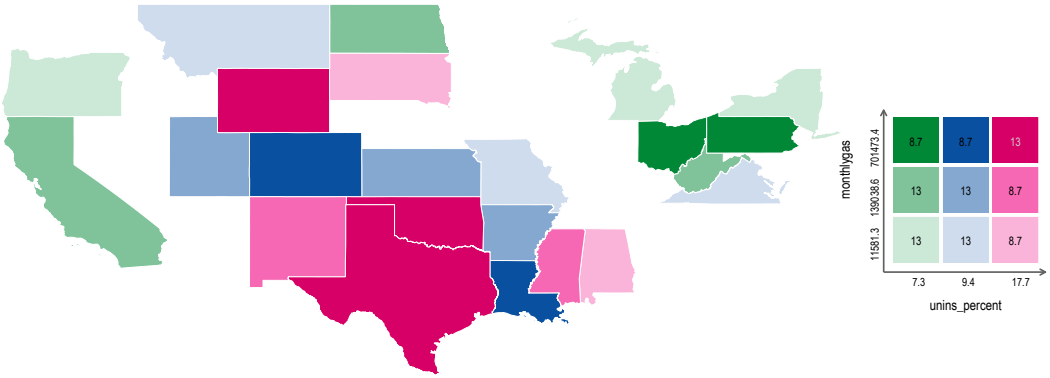
The following figure illustrates the association between state-level monthly NGD production in January 2018 and the percent of residents lacking health insurance in the producing states. Thanks to Asjad Naqvi for development and enhancement of `bimap`.

Those at greater risk to environmental burdens are often living in disadvantaged or minoritized communities. Some studies have found that infants born to Hispanic or Black women, women of lower socioeconomic circumstances, or women with lower educational attainmment had greater reductions in birth weight, and increased risk of LBW, SGA or preterm birth with NGD exposure than women in more favorable socioeconomic circumstances.

The following figure illustrates the association between state-level monthly NGD production in January 2018 and the percent of residents lacking health insurance in the producing states. Thanks to Asjad Naqvi for development and enhancement of `bimap`.

Natural gas production and percent lacking health insurance, 2018

Colors defined by tercile cutoffs of monthlygas and percent lacking health insurance



Empirical strategy

In order to conduct a more comprehensive study of these associations, we accessed microdata natality files provided by the National Center for Health Statistics for 2005–2018 with detailed demographic and health-related information on the mother and baby, with residence identified at the level of the US county.

This included self-reported race/ethnicity, education, age, nativity (US or foreign) and marital status at the time of delivery. Also, an indicator for prenatal smoking, parity (birth order), month of prenatal care intiation, infant sex and gestational age.

In the full analysis, outcome variables include birth weight, gestational age, and indicators of LBW (<2500g), preterm birth (<37 weeks), and SGA (< 10th percentile for gestational age and sex).

Empirical strategy

In order to conduct a more comprehensive study of these associations, we accessed microdata natality files provided by the National Center for Health Statistics for 2005–2018 with detailed demographic and health-related information on the mother and baby, with residence identified at the level of the US county.

This included self-reported race/ethnicity, education, age, nativity (US or foreign) and marital status at the time of delivery. Also, an indicator for prenatal smoking, parity (birth order), month of prenatal care intiation, infant sex and gestational age.

In the full analysis, outcome variables include birth weight, gestational age, and indicators of LBW (<2500g), preterm birth (<37 weeks), and SGA (< 10th percentile for gestational age and sex).

Empirical strategy

In order to conduct a more comprehensive study of these associations, we accessed microdata natality files provided by the National Center for Health Statistics for 2005–2018 with detailed demographic and health-related information on the mother and baby, with residence identified at the level of the US county.

This included self-reported race/ethnicity, education, age, nativity (US or foreign) and marital status at the time of delivery. Also, an indicator for prenatal smoking, parity (birth order), month of prenatal care intiation, infant sex and gestational age.

In the full analysis, outcome variables include birth weight, gestational age, and indicators of LBW ($<$2500g), preterm birth ($<$37 weeks), and SGA ($<$ 10th percentile for gestational age and sex).

The analytic sample included 33,859,409 singleton births with a gestational age of 30–44 weeks to women aged 16-49 years in the 28 US states with any NGD activity over the study period.

We acquired commercial monthly gas production data (million cubic feet, or MCF) from Enverus (enverus.com) for every well in the US for 2004–2018. The wells were geocoded and associated with the county, parish or borough in which they are located.

The analytic sample included 33,859,409 singleton births with a gestational age of 30–44 weeks to women aged 16-49 years in the 28 US states with any NGD activity over the study period.

We acquired commercial monthly gas production data (million cubic feet, or MCF) from Enverus (enverus.com) for every well in the US for 2004–2018. The wells were geocoded and associated with the county, parish or borough in which they are located.

To link these datasets, we used geocoded measures of county centroids, provided by the US Census Bureau, which identify the population-weighted central point in each county. The minimum distances between the centroid and each well were computed using trigonometry.

The county-level NGD production was weighted by location, using both inverse distance weighting and inverse distance squared weighting, allowing for a nonlinear relationship between distance and production.

The weighted production was summed at the county level and averaged over the estimated month of conception and the eight following months. The nine-month county averages were merged with the birth certificate data by month and year of birth.

To link these datasets, we used geocoded measures of county centroids, provided by the US Census Bureau, which identify the population-weighted central point in each county. The minimum distances between the centroid and each well were computed using trigonometry.

The county-level NGD production was weighted by location, using both inverse distance weighting and inverse distance squared weighting, allowing for a nonlinear relationship between distance and production.

The weighted production was summed at the county level and averaged over the estimated month of conception and the eight following months. The nine-month county averages were merged with the birth certificate data by month and year of birth.

To link these datasets, we used geocoded measures of county centroids, provided by the US Census Bureau, which identify the population-weighted central point in each county. The minimum distances between the centroid and each well were computed using trigonometry.

The county-level NGD production was weighted by location, using both inverse distance weighting and inverse distance squared weighting, allowing for a nonlinear relationship between distance and production.

The weighted production was summed at the county level and averaged over the estimated month of conception and the eight following months. The nine-month county averages were merged with the birth certificate data by month and year of birth.

Empirical analyses were carried out using linear regression for birth weight and gestational age and probit models for dichotomous outcomes.

Models were adjusted for the factors described earlier as well as birth year, with county random effects and clustered standard errors by county.

Additional models included interactions between NGD production and race/ethnicity, as Black and Hispanic women have been disproportionately affected by prenatal exposure to pollutants.

Empirical analyses were carried out using linear regression for birth weight and gestational age and probit models for dichotomous outcomes.

Models were adjusted for the factors described earlier as well as birth year, with county random effects and clustered standard errors by county.

Additional models included interactions between NGD production and race/ethnicity, as Black and Hispanic women have been disproportionately affected by prenatal exposure to pollutants.

Empirical analyses were carried out using linear regression for birth weight and gestational age and probit models for dichotomous outcomes.

Models were adjusted for the factors described earlier as well as birth year, with county random effects and clustered standard errors by county.

Additional models included interactions between NGD production and race/ethnicity, as Black and Hispanic women have been disproportionately affected by prenatal exposure to pollutants.

Computational challenges

Handling of the 33+ million birth certificate records for models with many covariates was carried out using Stata/MP 17 on 36-core and 48-core Linux clusters.

Larger computational challenges were encountered with the Enverus NGD data, which was provided as header files for each well and history files containing each well's monthly production history.

The number of wells per state ranged from 8 in Maryland and 37 in Oregon to 119,289 in Oklahoma and 233,882 in Texas. The Texas data contained 26.3 million well-month observations, and the full data set for all 28 states contained 128.3 million well-month observations.

Computational challenges

Handling of the 33+ million birth certificate records for models with many covariates was carried out using Stata/MP 17 on 36-core and 48-core Linux clusters.

Larger computational challenges were encountered with the Enverus NGD data, which was provided as header files for each well and history files containing each well's monthly production history.

The number of wells per state ranged from 8 in Maryland and 37 in Oregon to 119,289 in Oklahoma and 233,882 in Texas. The Texas data contained 26.3 million well-month observations, and the full data set for all 28 states contained 128.3 million well-month observations.

Computational challenges

Handling of the 33+ million birth certificate records for models with many covariates was carried out using Stata/MP 17 on 36-core and 48-core Linux clusters.

Larger computational challenges were encountered with the Enverus NGD data, which was provided as header files for each well and history files containing each well's monthly production history.

The number of wells per state ranged from 8 in Maryland and 37 in Oregon to 119,289 in Oklahoma and 233,882 in Texas. The Texas data contained 26.3 million well-month observations, and the full data set for all 28 states contained 128.3 million well-month observations.

The computation of time series county-level weighted production data was done separately for each state, using a custom Mata routine modified from the `compcty2` routine presented in *Introduction to Stata Programming, Second Edition* (Baum, 2016).

This routine takes as input the county (FIPS) code, latitude and longitude of the county centroid as identified by the Census Bureau, the latitude and longitude of the well and computes the distance, inverse distance and inverse squared distance for each well.

The computation of time series county-level weighted production data was done separately for each state, using a custom Mata routine modified from the `compcty2` routine presented in *Introduction to Stata Programming, Second Edition* (Baum, 2016).

This routine takes as input the county (FIPS) code, latitude and longitude of the county centroid as identified by the Census Bureau, the latitude and longitude of the well and computes the distance, inverse distance and inverse squared distance for each well.

Key findings

Across the 28 states, 53.8% had NGD production during the study period. Between 2005–2018, NGD production increased by 79%, with half of states decreasing production by 35% and the other half increasing production by 782%. Leading states' production increases:

- Ohio 2995%
- Pennsylvania 1090%
- Tennessee 1910%
- North Dakota 1382%
- West Virginia 717%

Key findings

Across the 28 states, 53.8% had NGD production during the study period. Between 2005–2018, NGD production increased by 79%, with half of states decreasing production by 35% and the other half increasing production by 782%. Leading states' production increases:

- Ohio 2995%
- Pennsylvania 1090%
- Tennessee 1910%
- North Dakota 1382%
- West Virginia 717%

Key findings

Across the 28 states, 53.8% had NGD production during the study period. Between 2005–2018, NGD production increased by 79%, with half of states decreasing production by 35% and the other half increasing production by 782%. Leading states' production increases:

- Ohio 2995%
- Pennsylvania 1090%
- Tennessee 1910%
- North Dakota 1382%
- West Virginia 717%

Key findings

Across the 28 states, 53.8% had NGD production during the study period. Between 2005–2018, NGD production increased by 79%, with half of states decreasing production by 35% and the other half increasing production by 782%. Leading states' production increases:

- Ohio 2995%
- Pennsylvania 1090%
- Tennessee 1910%
- North Dakota 1382%
- West Virginia 717%

Key findings

Across the 28 states, 53.8% had NGD production during the study period. Between 2005–2018, NGD production increased by 79%, with half of states decreasing production by 35% and the other half increasing production by 782%. Leading states' production increases:

- Ohio 2995%
- Pennsylvania 1090%
- Tennessee 1910%
- North Dakota 1382%
- West Virginia 717%

Overall, prenatal exposure to NGD production increased adverse birth outcomes.

We found that a 10% increase in NGD production in a county decreased mean birth weight by 1.48g (95% CI= -2.60, -0.37).

Although the overall effect of -1.48g is small, there were significant interactions of NGD production exposure with race/ethnicity.

A 10% increase in NGD production decreased birthweights of babies born to Black mothers by 10.19g (95% CI= -13.56, -6.81) and to Asian women by 2.76g (95% CI= -5.05, -0.46). No significant reductions were found for other racial/ethnic groups.

Increases in NGD production significantly raised the risk of LBW (low birth weight) or SGA (small for gestational age) infants, with small effects sizes. Significant interactions with race/ethnicity found higher risks for Black, Hispanic and Asian women.

Overall, prenatal exposure to NGD production increased adverse birth outcomes.

We found that a 10% increase in NGD production in a county decreased mean birth weight by 1.48g (95% CI= -2.60, -0.37).

Although the overall effect of -1.48g is small, there were significant interactions of NGD production exposure with race/ethnicity.

A 10% increase in NGD production decreased birthweights of babies born to Black mothers by 10.19g (95% CI= -13.56, -6.81) and to Asian women by 2.76g (95% CI= -5.05, -0.46). No significant reductions were found for other racial/ethnic groups.

Increases in NGD production significantly raised the risk of LBW (low birth weight) or SGA (small for gestational age) infants, with small effects sizes. Significant interactions with race/ethnicity found higher risks for Black, Hispanic and Asian women.

Overall, prenatal exposure to NGD production increased adverse birth outcomes.

We found that a 10% increase in NGD production in a county decreased mean birth weight by 1.48g (95% CI= -2.60, -0.37).

Although the overall effect of -1.48g is small, there were significant interactions of NGD production exposure with race/ethnicity.

A 10% increase in NGD production decreased birthweights of babies born to Black mothers by 10.19g (95% CI= -13.56, -6.81) and to Asian women by 2.76g (95% CI= -5.05, -0.46). No significant reductions were found for other racial/ethnic groups.

Increases in NGD production significantly raised the risk of LBW (low birth weight) or SGA (small for gestational age) infants, with small effects sizes. Significant interactions with race/ethnicity found higher risks for Black, Hispanic and Asian women.

Overall, prenatal exposure to NGD production increased adverse birth outcomes.

We found that a 10% increase in NGD production in a county decreased mean birth weight by 1.48g (95% CI= -2.60, -0.37).

Although the overall effect of -1.48g is small, there were significant interactions of NGD production exposure with race/ethnicity.

A 10% increase in NGD production decreased birthweights of babies born to Black mothers by 10.19g (95% CI= -13.56, -6.81) and to Asian women by 2.76g (95% CI= -5.05, -0.46). No significant reductions were found for other racial/ethnic groups.

Increases in NGD production significantly raised the risk of LBW (low birth weight) or SGA (small for gestational age) infants, with small effects sizes. Significant interactions with race/ethnicity found higher risks for Black, Hispanic and Asian women.

Overall, prenatal exposure to NGD production increased adverse birth outcomes.

We found that a 10% increase in NGD production in a county decreased mean birth weight by 1.48g (95% CI= -2.60, -0.37).

Although the overall effect of -1.48g is small, there were significant interactions of NGD production exposure with race/ethnicity.

A 10% increase in NGD production decreased birthweights of babies born to Black mothers by 10.19g (95% CI= -13.56, -6.81) and to Asian women by 2.76g (95% CI= -5.05, -0.46). No significant reductions were found for other racial/ethnic groups.

Increases in NGD production significantly raised the risk of LBW (low birth weight) or SGA (small for gestational age) infants, with small effects sizes. Significant interactions with race/ethnicity found higher risks for Black, Hispanic and Asian women.

Discussion

As NGD production has increased across the US, it has imperiled infant health, with greater effects in infants born to minoritized women. Although those women do not live in counties with higher NGD production than white women, they may be subject to other socioeconomic factors, such as restrictions on Medicaid coverage.

Data limitations on the birth certificate only allow access to the geocoding of mothers at the county level, while some previous studies have had access to more granular information for limited samples. Some studies have focused on well counts rather than production volume, which is essential to gauge exposure.

Discussion

As NGD production has increased across the US, it has imperiled infant health, with greater effects in infants born to minoritized women. Although those women do not live in counties with higher NGD production than white women, they may be subject to other socioeconomic factors, such as restrictions on Medicaid coverage.

Data limitations on the birth certificate only allow access to the geocoding of mothers at the county level, while some previous studies have had access to more granular information for limited samples. Some studies have focused on well counts rather than production volume, which is essential to gauge exposure.

These results contribute to the growing body of research on the impact of natural gas and oil production on infant health. These activities consume fresh water, generate wastewater containing many chemicals, require diesel engines, and produce traffic volume to transport sand and chemicals.

The pollutants generated from these activities infiltrate the air and water of surrounding communities which can be ingested by pregnant women and cross the placenta, resulting in disruptions to fetal growth.

These results contribute to the growing body of research on the impact of natural gas and oil production on infant health. These activities consume fresh water, generate wastewater containing many chemicals, require diesel engines, and produce traffic volume to transport sand and chemicals.

The pollutants generated from these activities infiltrate the air and water of surrounding communities which can be ingested by pregnant women and cross the placenta, resulting in disruptions to fetal growth.

Although there are economic benefits to domestic natural gas and oil production, there are also increasing concerns about fossil fuel production as worldwide concerns over climate change press for reductions in nonrenewable energy sources.

The environmental and health risks associated with fracking have led seven US states (Maryland, Vermont, Washington, Pennsylvania, New Jersey, Delaware and New York) and many European countries to ban fracking in more recent years. The findings of this study suggest that this is appropriate public policy from an environmental and health standpoint.

Although there are economic benefits to domestic natural gas and oil production, there are also increasing concerns about fossil fuel production as worldwide concerns over climate change press for reductions in nonrenewable energy sources.

The environmental and health risks associated with fracking have led seven US states (Maryland, Vermont, Washington, Pennsylvania, New Jersey, Delaware and New York) and many European countries to ban fracking in more recent years. The findings of this study suggest that this is appropriate public policy from an environmental and health standpoint.