Investigation On The Association Between Unconventional Oil And Gas Development And Traffic Accident Rates In Ohio

Hubian Wang
1178176923@qq.com

Follow this and additional works at: https://elischolar.library.yale.edu/ysphtdl

Recommended Citation
https://elischolar.library.yale.edu/ysphtdl/2107

This Open Access Thesis is brought to you for free and open access by the School of Public Health at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Public Health Theses by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.
Investigation on the association between unconventional oil and gas development and traffic accident rates in Ohio

Name: Hubian Wang

Year Completed: 2021

Year Degree Awarded: 2021

Degree Awarded: Master of Public Health

Department: School of Public Health

Advisor/Committee Chair: Nicole Deziel

Committee Members: Kai Chen
Abstract

The oil and gas drilling activities had greatly developed in Ohio in past decade. Unconventional natural gas development is a newly developed technology, which requires more heavy-truck trips in the drilling process. Limited studies investigated the association between drilling activities and increased rate of traffic accidents.

Ecological study was conducted to analyze the association. Number of unconventional wells were obtained as exposure variable from Ohio department of natural source, and the data of traffic crashes were collected as outcome variable from Ohio department of public safety. Other variables including sociodemographic data and spatial distribution of primary roads were examining as confounders. Poisson regression models were used to conduct multivariable regression analysis.

Continuous, binary and categorical variables were defined to indicate the exposure. The fitting result showed that the traffic incident rate ratio estimate for well number was 0.9944 (0.9939, 0.9949) for continuous model (per 10 wells), 0.9247 (0.9210, 0.9284) for binary model (drilled vs not drilled), 0.9460 (0.9417, 0.9503) and 0.8629 (0.8561, 0.8699) for categorical model (medium vs low drilling intensity, high vs low drilling intensity, respectively). The study also found counties with primary roads across had higher traffic incident rate compared to counties without primary roads across.

Key words: environmental health science, unconventional oil and gas development, traffic incident rate
Acknowledgement

I would like to express my sincere gratitude to my thesis advisor Dr. Nicole Deziel for her guidance. I learned a lot from Dr. Deziel during the thesis, including discovering public health problems, obtaining data, data analysis and using epidemiological tools to address public health problems. Dr. Deziel pointed out the core direction for me and encouraged me all the time, her help was necessary for me to finish the thesis.

I would also like to express my sincere gratitude to my thesis reader Dr. Kai Chen. Dr. Chen provided me very helpful and rigorous advice, and I learned useful data analysis knowledge from previous course taught by Dr. Chen.

In addition, I would like to thank Chao Wu for his suggestion in our discussion and help about my thesis format. Finally, I would like to thank my parents and friends for their encouragement and support all the time.
Table of content

1. Introduction ......................................................................................................................... 1

2. Methods ................................................................................................................................. 4
   2.1 Study design ...................................................................................................................... 4
   2.2 Data sources ..................................................................................................................... 4
       2.2.1 Traffic data ............................................................................................................... 4
       2.2.2 Horizontal oil and gas wells .................................................................................... 5
       2.2.3 Demographic data ................................................................................................... 6
       2.2.4 Primary roads data .................................................................................................. 8
   2.3 Analysis methods ............................................................................................................. 9

3. Results ................................................................................................................................... 12
   3.1. Study area description .................................................................................................... 12
   3.2 Regression analysis result ............................................................................................... 15

4. Discussion ............................................................................................................................. 21

Tables ....................................................................................................................................... 28
Figures ....................................................................................................................................... 34
Reference ................................................................................................................................... 39
List of tables

Table 1. Distribution of sociodemographic factors among ever drilled and not drilled counties in Ohio in predrilling period (2004–2009). ................................................................. 28

Table 2. Distribution of sociodemographic factors among ever drilled and not drilled counties in Ohio in predrilling period (2010–2017). ................................................................. 28

Table 3. Rate ratios estimate from crude regression models for association between drilling activity and traffic incidents in Ohio 2010-2017 (n=704 county-years) ........................................ 29

Table 4. Rate ratios estimate from multivariable regression models for association between drilling activity and traffic incidents in Ohio 2010-2017 (n=704 county-years) .................... 30

Table 5. Rates ratios estimate from multivariable regression model by different incident types in Ohio 2010-2017 (n=704 county-years) ................................................................. 30

Table 6. Measurement of goodness of matching .................................................................. 31

Table 7. Rate ratios estimate from multivariable regression models among matched counties in Ohio 2010-2017 (n=384 county-years) ................................................................. 32

Table 8. Rate ratios estimate from multivariable regression models using continuous road variable in Ohio 2010-2017 (n=704 county-years) ................................................................. 32

List of figures

Figure 1. Total horizontal well numbers by Ohio counties .................................................. 34

Figure 2. Mean incident rates across counties and time by drilling activity ..................... 35

Figure 3. Total incident rates of 2010 to 2017 by Ohio counties ...................................... 36

Figure 4. Result of Spearman correlation analysis of demographic variables ............... 37

Figure 5. Result of matched counties in Ohio ................................................................. 38
1. Introduction

Natural gas has been regarded as a clean and efficient energy source compared to coal under climate change, and was encouraged to develop to protect environment and reduce air pollutants and greenhouse gas emission alternative to traditional coal burning. The United States holds a large quantity of oil and natural gas resources. The oil and gas drilling industry has developed rapidly in the state of Ohio based on the Marcellus and Utica shale. Up to November 2020, the new-well gas production per rig of Marcellus Shale reached 26,700 thousand cubic feet per day, while the number was 14,850 in November 2015. Unconventional natural gas development is a newly developed drilling technology, which includes high-volume horizontal hydraulic fracturing in the drilling process. Ohio drilled the first unconventional horizontal shale wells in 2010-2011, and the Division of Oil and Gas Resources Management has been regulated the drilling activities and issued drilling permits for wells.

The development of oil and gas industries could lead to economic benefits, including increased employment and household income. However, the development could raise environmental and public health concerns in the communities where well sites was located. Studies showed that the drilling activities might lead to the excess level of contaminants in drinking water and air pollutants near drilling sites, including but not limited to adverse birth outcomes, cancer incidence and sexually transmitted
infections\textsuperscript{12}. Many studies have investigated the environmental exposures and the pathways of chemical and analyzed their public health impact\textsuperscript{3, 11}, but limited study have researched the outcome of increasing traffic accidents by increased drilling wells\textsuperscript{13}.

The unconventional drilling process requires high volumes of truck traffic to fulfill the transportation requirement\textsuperscript{14}. Heavy-truck trips are greatly needed to transport the construction material, saltwater, and backflow water that are necessary in the processes\textsuperscript{15, 16}. It is estimated that the construction and development of each well in Marcellus would require 1500 heavy-truck trips\textsuperscript{13}. The development of gas well could potentially lead to increased rates of traffic accidents. The increased volume of heavy-truck trips could exceed the designed capacity of roads, which means bringing more burden to transportation infrastructure\textsuperscript{17, 18}. The vehicles on roads may become denser and the roads may suffer from degradation and finally results in a higher risk of traffic accidents\textsuperscript{19}. In addition, drivers work for fracking industries are permitted to drive for longer hours, which means there is a higher risk of fatigue driving\textsuperscript{20}. All these factors making the development of fracking potentially be associated with increased risk of traffic accidents.

Previous studies reported that the rate ratio of truck traffic accidents was 1.07 in counties with more drilling activity compared to less drilling activity in Colorado\textsuperscript{21}. Another study concluded the rates of vehicle crashes in heavily drilled counties was 15-23 % higher than not drilled control counties in 2010-2012 and rates of heavy truck
crashes was 61-65% higher in 2010-2011 in the state of Pennsylvania\textsuperscript{13}. The severe injury truck crashes increased more than 12 times during 2008-2012 in drilling area of North Dakota, while the other part only had increased less than 2 times\textsuperscript{22}. Ohio has also been experienced development of the unconventional oil and gas industry, yet the study researching the effects on traffic was still limited. It is necessary to address the public health issue of traffic incidents and investigate the association between traffic incident rates and unconventional drilling activities.
2. Methods

2.1 Study design

The study was designed to research the relationship between increasing unconventional drilling activities and traffic accident rates from 2004 through 2017. The study population is designed to be all people and drivers using road or highway transport in the state of Ohio. An ecological study was designed to be the major part of the investigation. The reason for selecting the ecological study is that the researched independent variables and outcome variable were described at the county level\(^2\). It is not appropriate to regard the unconventional drilling activities as an exposure to individuals, because each drilling activity was always affecting a neighborhood of the drilling location\(^5\). In addition, the result of ecological analysis could provide reasonable advice on the level of policymaking, such as taking additional safety measures or making special traffic notices on the heavily drilling counties\(^1\). Therefore, the ecological study method might be the optimal selection for this study. County-year was defined as the unit of observation in the ecological study. There are 88 counties in total and the study time period was 14 years, including baseline predrilling period and unconventional drilling period, thus there were totally 1232 observations in the ecological study.

2.2 Data sources

2.2.1 Traffic data
The annual traffic data contains the total number of all kinds of crashes in each county, including fatal crashes, injury crashes and properties damaged only (PDO) crashes. The annual traffic datasets were collected from the Ohio traffic crash facts books, which were published annually by the Ohio State Highway Patrol division of the Ohio Department of Public Safety. However, the traffic crash data by vehicle type was not available through the public source, so the data contained crashes related to heavy trucks and other vehicle types. The table 7.01 Total Crashes by County from the Ohio traffic crash facts books was the raw data source for the annual traffic data in this study. The crash rate of each county-year was calculated by dividing the annual total crashes by estimated population of that county, while the estimated county population might slightly differ in each year.

2.2.2 Horizontal oil and gas wells

The number of annually drilled horizontal oil and gas wells in each county were collected for analysis in this study. The raw data was obtained from The Division of Oil and Gas Resources Management of the Ohio Department of Natural Resource. The division integrated the weekly data of shale activity until March 2021, and the reports contained the time spot when the permit issued and the location of each horizontal well in Utica and Marcellus Shale. The total numbers of annually drilled wells located in each shale were calculated in county level. The first unconventional horizontal shale wells were drilled in Ohio in 2010-2011, thus the number of horizontal wells in each county was zero in the period of 2004 to 2009.
The variable of total horizontal well numbers for each county-year observation in this study was defined as the accumulated numbers of horizontal wells in each county until the year observed, which was calculated by adding up all the numbers of annually drilled wells in previous and observed years in that county. Using the accumulated number was based on the consideration of the lifetime of wells, which might exceed 1500 days, and the truck transportation might still be active after drilling period\textsuperscript{13, 25}. Based on the total horizontal well numbers, all counties in Ohio were divided into ever-drilled and non-drilled groups, and a dummy variable named ever-drilled was created to indicate it. The counties with total horizontal well numbers being equal to zero in all observed years would have ever-drilled coded as zero and other counties were coded as one. Based on this definition, the ever-drilled variable stayed constant for the same county across different years. In addition, a categorical variable named drilled intensity was created for all county-years. A cutoff value of 20 was selected since a previous study reported the effect was observed only above a threshold of 20 new wells. Therefore, the county-years that had total horizontal well numbers equal to zero were defined as low intensity, and the county-years had total horizontal well numbers in range of 0 to 20 were defined as medium intensity while those county-years had total horizontal well numbers greater than 20 were defined as high intensity. Thus, a county might be categorized into different drilled intensity in different periods, which was different from the ever-drilled variable.

\textbf{2.2.3 Demographic data}
The demographic data being researched in this study included estimated population, percent of white people, percent of black people, percent of Hispanic people, percent of males, percent of females, percent of people having health insurance, percent of population between 15 to 29 years old, percent of people having high school degree, percent of people having bachelor’s degree, median household income. All these demographic variables were continuous and at the county level for each year. According to the Ohio Traffic Facts Books of 2017 published by Ohio Traffic Safety Office, the 16 to 30 years old contributed the largest part of persons killed or injured in traffic accidents among all age groups. Male deaths were more than two times female deaths. The distribution of age group and sex among victims was consistent with the distribution among drivers, indicating age and sex may impact traffic accidents occurrence. Other variables, including education and income-related variables may affect traffic incident rates because drivers in low education and income level tend to have more risky behaviors.

The data of estimated population and percent of different races were obtained from Annual County Resident Population Estimates by Age, Sex, Race, and Hispanic Origin dataset published by the U.S. Census Bureau Population Division. The dataset was released in 2020 and contained the annual estimation of population at the county level based on the 2010 Census population data.

The data of percent of sex and population between 15 to 29 years old were obtained
from the Annual County and Puerto Rico Municipio Resident Population Estimates by
Selected Age Groups and Sex dataset\textsuperscript{28}. This dataset was also published by the U.S.
Census Bureau Population Division in 2020 and based on the 2010 Census data.

The data of percent of people having high school and bachelor’s degree were obtained
from the American Community Survey (ACS)\textsuperscript{29}, social characteristics subject. The data
of percent of people having health insurance and median household income level were
obtained from the ACS, economic characteristics subject\textsuperscript{29}. The 2019 data was the latest
version of ACS results, and the data for previous years in this study period can be
accessed in the data profile.

\textbf{2.2.4 Primary roads data}

The data of primary roads and county boundaries of Ohio were collected from US
Census Bureau in the format of shapefile\textsuperscript{30}. A binary variable Road across and a
continuous variable Distance were created as indicator of primary road variables. Road
across was defined as whether there were any primary roads across the county. The
nearest distance between primary roads and the polygon representing the county was
calculated, and the nearest distance equal to 0 means there was at least one primary
road across the county. Distance was defined as the nearest distance between primary
roads and the centroid of the polygon representing the county, and the unit was 10 km.
Road across was used in multivariable regression model analysis and Distance was used
in sensitivity analysis.
2.3 Analysis methods

The study investigated the rate ratios and 95% confidence intervals for the relationship between traffic accidents and unconventional oil and gas activities by conducting regression analyses. A generalized linear model was used to conduct Poisson regression analysis. The numbers of crashes was the primary outcome variable in the study, so the Poisson regression is the most widely selected choice for analysis in this situation\textsuperscript{31}. In the Poisson regression model, the logarithm of estimates of population in each county-year was set as an offset term to compare the rate ratio of crashes.

Unconventional drilling activities were the major independent variable of interest for all analysis models. It was described as binary, categorical or continuous variables in different models. In the binary variable analysis, whether a county was ever drilled (yes or no) became the unconventional drilling activities indicator. In the categorical variable analysis, the three-categorical drilling intensity (low, medium or high) was the exposure variable. In the continuous variable analysis, the total number of horizontal wells was included in the model.

The correlation analysis was conducted to check whether collinearity appeared among all continuous covariates, including percent of white people, percent of black people, percent of Hispanic people, percent of males, percent of females, percent of people having health insurance, percent of population between 15 to 29 years old, percent of people having high school degree, percent of people having bachelor’s degree, median
household income. A Spearman correlation test was conducted among these variables and a threshold of 0.7 was set for $|\rho_{\text{Spearman}}|$ to be considered as highly correlated\textsuperscript{12}. If two covariates were highly correlated then one of them would be excluded in the regression analysis. The statistical significance level was set to be 0.05 in regression analysis.

Experiences showed that the traffic crash rates vary in different counties and might be more or less influenced by uncertain factors, which were heterogeneous among counties. Therefore, a matching process was conducted to minimize the effect of bias in the estimation of association\textsuperscript{13}. Ever drilled counties were defined as treated subjects and not drilled counties were defined as untreated subjects in this study. The closeness of matching was defined as Mahalanobis distance calculated using the linear propensity score and two key variables: the population and crash rate in 2009 of each county, indicating the baseline value for the pre-drilling period. The caliper, defined as restriction on the distance between the untreated and treated subjects, was set to be 0.25 times the pooled estimate of the standard deviation of the logits of the propensity score across treated and untreated subjects\textsuperscript{32}. The structure of the matching was greedy and 1-to-1 matching by setting random order for observations, which means searching an untreated subject with the nearest closeness to the randomly-ordered treated subject within the caliper, and one untreated subject was matched to one treated subject. Standardized difference and variance ratio for key weighted variables were calculated
to measure the goodness of matching, and a standardized difference <0.10 and a variance ratio between 0.8 to 1.25 were considered as ideally balanced matching\textsuperscript{33-35}. Then the regression analyses were conducted within the matched groups to reduce the effect of uncertain factors.

The analysis processed were conducted by the software of SAS 9.4 (SAS Institute, Cary, NC, USA) and R (version 4.0.3). The figures of maps were plotted by the software of ArcGIS Pro (Esri). The geometric calculation were also conducted by the software of ArcGIS Pro (Esri).
3. Results

3.1. Study area description

The distribution of sociodemographic factors in the predrilling period (2004-2009) among ever drilled and not drilled counties were described in Table 1 and the distribution in the drilling period (2010-2017) were described in Table 2. According to the definition mentioned above, 24 counties were defined as ever drilled counties and 64 counties were defined as not drilled. Student’s t-test was conducted on each sociodemographic factor between the two groups under a significance level of 0.05.

According to the result from Table 1, the mean difference of all sociodemographic factors between ever drilled and not drilled counties were statistically significant, except the variable of percent of males. During the period of 2004 to 2009, the ever drilled counties had a smaller population, higher percent of white, lower percent of black, Hispanic, population had insurance, population aged 15-29 years old, population had high school degree and bachelor’s degree, and lower median income level when compared to not drilled counties. Percent of males were similar between the two groups. Generally speaking, those ever drilled counties were less developed than not drilled counties considering the variables indicating insurance, education and income.

Similar conclusion can be reached according to the result of Table 2. The mean difference of all sociodemographic factors between ever drilled and not drilled counties were also statistically
significant, except the variable of percent of males. During the period of 2010 to 2017, the ever drilled counties had a smaller population, higher percent of white, lower percent of black, Hispanic, population had insurance, population aged 15-29 years old, population had high school degree and bachelor’s degree, and lower median income level when compared to not drilled counties. Percent of males were similar between two groups. Those ever drilled counties were still less developed than not drilled counties considering the variables mentioned above. However, the values of these variables became larger compared to the previous period, indicating developments in both non drilled and ever drilled counties.

The total horizontal well numbers of each county were showed in Figure 1. The number in the plot indicated the accumulated number of drilled horizontal wells in each county until 2017. The drilled counties were distributed in the eastern part of Ohio, which was consistent with the location of Marcellus and Utica shale, according to the information from the department of natural source. Columbiana, Carroll, Jefferson, Harrison, Guernsey, Belmont, Noble and Monroe were the counties with the highest drilling intensity among all counties, and all of them were located in the eastern most part of Ohio.

The mean incident rates for different kinds of crashes across counties and time by drilling activity were described in Figure 2. Total crash rate, injury crash rate and PDO crash rate all had a generally decreasing trend from 2004 to 2013, followed by a smaller increase after 2013. Not drilled counties had similar trend of total crash rate, injury rate and PDO crash rate with total counties, and the rates were higher than ever drilled
counties in almost all observed years. For the fatal crash rate, the trend changed more intensely across years, which might due to a much smaller baseline value (1.13 to 1.63 per 10000 population) compared to other three kinds of rates. The fatal crash rate had an unpredictable trend across years and sometimes the rate of ever drilled counties were larger than not drilled counties (in the year of 2012 and 2014). Generally speaking, the mean crash rates were decreasing over time, and the mean rate of total crashes, fatal crashes, injury crashes and PDO crashes among ever drilled counties were less than mean rate among total counties and not drilled counties, respectively.

The traffic incident rates by county were described in Figure 3. The incident rates in Figure 3 was calculated by the accumulated total crashes in 2010 to 2017 divided by total person-years in 2010 to 2017, and were categorized into 5 quantiles, the darker color means higher incident rates of each county. The figure also plotted the primary roads and metropolitans with population larger than 1,000,000 in Ohio. There were six metropolitans in Ohio, Columbus (Franklin county), Akron (Summit county), Cincinnati, (Hamilton county), Cleveland (Cuyahoga county), Toledo (Lucas county) and Dayton (Montgomery county), respectively. The figure showed that the incident rates were obviously higher in the counties where metropolitans were located or primary roads were near. Most counties with incident rates in quantile 1 were located in eastern and southeastern region of the state when referring to Figure 1 at the same time. In the drilled region, there were not many primary roads that went through the
counties, and some counties had incident rates in the lowest quantile.

### 3.2 Regression analysis result

Table 3 presented the regression coefficient estimates from the crude analysis. The Poisson regression model used in this crude analysis can be described as following equations:

\[
\log \left( \frac{\text{Total crashes}}{\text{population}} \right) = \beta_0 + \beta_1 \times \text{Well number variable}
\]

Where well number variable refers to total horizontal wells (continuous) or ever drilled (binary) or drilling intensity (3 categories). The model used population as an offset term. The observations in this analysis included all county-years in the drilling period (2010-2017).

The result of continuous model indicated that the mean traffic incident rate became 0.9992 times when there is 1 total horizontal well increased. The result of binary model showed that the mean traffic incident rate of ever drilled counties were 0.8966 times compared to the not drilled counties. The result of categorical model showed that the mean traffic incident rate of counties with medium drilling intensity was 0.9114 times compared to the counties with low drilling intensity; the mean traffic incident rate of counties with high drilling intensity was 0.8514 times compared to the counties with low drilling intensity. The baseline rate estimates from the three models were consistent with each other, and all estimates from all the crude models were statistically significant.
Correlation test was conducted prior to the multiple variables regression analysis. Figure 4 showed the result of Spearman correlation test. Under the threshold of 0.7, the variables of percent of white people and percent of black people, percent of people having high school degree and percent of people having health insurance, percent of people having bachelor’s degree and percent of people having high school degree, median household income and percent of people having high school degree were considered highly correlated.

Therefore, the adjusted model included well number variables, road across, percent of white people, percent of Hispanic people, percent of males, percent of people having health insurance, percent of population between 15 to 29 years old, percent of people having bachelor’s degree, median household income. The Poisson regression model used in multivariable regression analysis can be described as following equations:

\[
\log\left( \frac{\text{TotalCrashes}}{\text{population}} \right) = \beta_0 + \beta_1 \times \text{Well number variable} + \beta_2 \times \text{percent white} + \beta_3 \times \text{percent hispanic} + \beta_4 \times \text{percent males} + \beta_5 \times \text{percent health insurance} + \beta_6 \times \text{percent population 15 – 29 years} + \beta_7 \times \text{percent bachelors degree} + \beta_8 \times \text{median household income} + \beta_9 \times \text{road across}
\]

Table 4 described the result of multiple variable regression between rates of total crashes and drilling activity and independent variables.

The result of the continuous model indicated that the mean traffic incident rate became 0.9944 times when there is ten total horizontal well increased and other variables are
fixed. The result of the binary model showed that the mean traffic incident rate of ever drilled counties was 0.9247 times compared to the not drilled counties when other variables are fixed. The result of categorical model showed that the mean traffic incident rate of counties with medium and high drilling intensity was 0.9460 and 0.8629 times compared to the counties with low drilling intensity when other variables are fixed, respectively. Compared with crude models, the rate ratio estimates for well number variable were larger and closer to 1 among all models, respectively. All estimates from all the multivariable models were statistically significant except percent of population between 15 to 29 years old in continuous model. The mean crash rates of counties with primary road across was 1.1270, 1.1512, 1.1368 times compared to the counties without primary road across respectively for three models. This result showed that whether there was primary road across the county was an important confounder with relatively large magnitude. Increase in percent of Hispanic people, percent of population between 15 to 29 years old, and median household income would lead to decrease in traffic incident rates. Increase in percent of males, percent of people having health insurance and percent of people having bachelor's degree would lead to increase in traffic incident rates.

Table 5 described the estimates for effect of well number variable and road across as major confounder on different types of incident rates, including fatal crashes, injury crashes and PDO crashes. All these models were adjusted for the variables described in
Table 4. According to the result from the continuous model, when the well number increases, the fatal crashes and injury crashes will contribute more to total crashes. The estimate for effect on fatal crashes was not significant. However, the binary model showed that there were larger rate ratios for fatal and injury crashes than total crashes among ever drilled counties compared to not drilled counties. The categorical model had a different result. When compared to low drilling intensity counties, the medium drilling intensity counties had larger rate ratio on injury crashes than total crashes, while high drilling intensity counties had larger rate ratios on fatal and injury crashes than total crashes. For injury and PDO crashes, the estimate effect of road across had consistent direction with total crashes, but the effect was inverse to fatal crashes. Counties with primary road across tend to had lower fatal crash rates compared to counties without primary road across. Although the estimate effect was not significant in binary and categorical model, the magnitude cannot be ignored compared to well number variable.

Figure 5 showed the matching result of counties in Ohio. The matched pairs were plotted by the same filling color and different border color. There were 24 matched pairs and therefore 48 counties were included in the matching analysis. There were no metropolitans located in matched counties, indicating the great difference between metropolitans and drilling counties.

The counties were matched based on crash rates and population in 2009 as the baseline,
Table 6 described the goodness of matching. The p-value was much larger than 0.05, indicating the difference of crash rate and population between drilled counties and not drilled counties were not significant. The standardized difference after matching was less than 0.10 and the variance ratio after matching was between 0.80-1.25, which provided evidence that the baseline crash rate and population were ideally balanced between drilled and not drilled counties after matching.

Table 7 presents the result of multivariable regression focusing on the drilling period among matched counties. The result was interesting when compared with the result based on all counties showed in Table 4. The coefficient estimates on well number variable were all larger than previous result, indicating the rate ratios of traffic accidents were relatively larger when drilling activities increased, although the associations were still slightly negative. The effect of percent of male and percent of population between 15-29 years old were inversed when compared with the previous result, which were negatively and positively associated with traffic accident rates, respectively. The association of age became more reasonable and consistent with previous study that young people are more likely to drive fast\textsuperscript{36, 37}. The effect of percent of population with health insurance, percent of population having bachelor’s degree and median household income level remained same direction, but all of them became towards null (rate ratios closer to 1) in this analysis.

The models were reran using distance to replace the road across variable for sensitivity
analysis. Other variables were same as previous models. Table 8 described the result of rate ratios estimate from the revised multivariable regression models. The distance variable had negative effect on traffic incident rates, the traffic incident rate would be 0.9736 time per 10 km increase of distance when other factors were fixed based on the continuous model, and 0.9962 and 0.9687 times for binary and categorical model, respectively. The counties without primary road across would had larger distance to the nearest primary road, so there would be inverse effect for distance to road across. Compared to Table 4, the estimate rate ratios for almost all variables stayed stable. The direction of estimated effects were same as results in Table 4 and the magnitudes were very similar for sociodemographic variables.
4. Discussion

The study investigated the traffic incident rates in Ohio where unconventional drilling activities were greatly developed since 2010 by county level ecological analysis. The Poisson regression model used crash counts as outcome variable and population as an offset, and variables including road across, percent of white people, percent of Hispanic people, percent of males, percent of people having health insurance, percent of population between 15 to 29 years old, percent of people having bachelor’s degree, median household income were controlled. The regression result showed that the mean traffic incident rate became 0.9944 times when there is 10 total horizontal well increased; the mean traffic incident rate of ever drilled counties were 0.9247 times compared to the not drilled counties; the mean traffic incident rate of counties with medium and high drilling intensity was 0.9460 and 0.8629 times compared to the counties with low drilling intensity when other variables are fixed, respectively. The estimates from the multivariable regression models were all statistically significant, and all showed negative association between drilling activities and traffic incident rates. When comparing with the crude models, the rate ratios estimates for well number variable were larger and closer to 1 among all models, respectively. The overall negative association also occurred when examining the rates according to traffic incident types. When examining the fatal crashes, the result showed that counties with primary road across tend to had lower fatal crash rates compared to counties without primary road.
across. The result of matching analysis presented less negative association between drilling activities and traffic incident rates, while some other variables showed inverse direction of association compared with results of multivariable regression analysis.

From the regression analysis above, the results from different models showed that the association estimates were generally consistent with each other, that the increase in drilling activities was slightly but negatively associated with traffic incident rates in observed county-years. The results were similar when examining different types of traffic crashes. Therefore, it can be inferred that the county-year with higher drilling activity intensity may had a relatively lower traffic incident rate compared to the county-year with lower drilling intensity during the research period in Ohio. The association was inverse compared to the expectation, and there may be several possible reasons. The counties with primary roads crossed by would generally had higher traffic incident rates, possibly due to the higher traffic volume and density. There were no primary roads across the counties that had most total horizontal wells (region colored by the darkest blue in Figure 1), and only a few counties with medium number of total horizontal wells had primary roads going through (region colored by lighter blue in Figure 1). The geographical deployment of primary roads limited the traffic volumes in the drilled counties, especially those with the highest drilling intensity. The counties where metropolitans located in also had higher traffic incident rates. The metropolitans had larger population, number of vehicles and better road infrastructure,
resulting in crowded traffic situations. The major metropolitans in Ohio were all located in not drilled counties, and contributed a large proportion of the higher traffic incident rates among not drilled counties. When combining the possible reasons above, another possible reason can be inferred. Drilling related truck drivers might pay more caution when driving in drilled counties with worse road conditions, but they might be less cautious and more tired when driving in primary roads afterwards. The trucks left drilled counties, and the higher traffic volume or other factors led to the occurrence of traffic accident to the trucks when they were located in the primary roads of not drilled counties, and the crashes were accounted into the statistics of not drilled counties.

The result of this study was inconsistent with some other previous studies. A study reported that the rates of vehicle crashes in heavily drilled counties was 15-23 % higher than not drilled control counties in 2010-2012 in the state of Pennsylvania\textsuperscript{13}, and rates of heavy truck crashes was 61-65% higher in 2010-2011. The time of observation in the Pennsylvania study was 2005-2012, while the unconventional oil and gas development was not started until 2010 in Ohio, therefore there was just a limited overlap proportion of the study time of these two studies. The outcome of interest in Pennsylvania study was crashes per million vehicle-miles per months, and this study used county-months as unit of observation. The study also considered the effect of 3-months window which was the expected time period of well construction and development. These differences may lead to the inconsistency between two studies.
Another study reported that the rate ratio of truck traffic accidents was 1.07 in counties with more drilling activity compared to less drilling activity in Colorado by county-level analysis. The study also concluded that number of homes and/or wells were positively associated with probability of multivehicle truck accidents with an injury by grid analysis. The Colorado study researched the incidence of truck accidents on a per capita basis in 2005-2013. However, the Colorado study did not include variables except population in their county-level analysis, and they only considered the number of homes in their grid analysis. The study contained all wells including conventional and unconventional drilled. The effects of all these factors can possibly explain the heterogeneity between the results.

The main strength of this study is the complete data and geographical variables considered in the regression analysis. The study investigated the detailed traffic data for long period (2004-2017), and included multiple sociodemographic variables as covariates in the regression analysis. By conducting multivariable regression analysis, more factors that were associated with traffic incident rates could be revealed, and the confounders affecting crash rates can be controlled. Considering the geographical variables in the analysis also provided evidence for the significant effect of heterogeneously distributed factors. The study also used several types of variables as the indicator of drilling activity intensity, partially reduced the problem of well number threshold for continuous variable and information loss for binary and categorical
variables. The rate ratio estimate from binary and categorical models had larger magnitude compared to continuous variable in the study, and the estimates differed among categories in the categorical model. The categorical well number variable could be considered the best to capture the exposure. The number of wells had huge discrepancy between different counties (0-530) that should not be ignored, and the different result among categories also confirmed this point. This county-level ecological study was also beneficial for community and policymaking. The results from the study showed that the traffic incident rates were also associated with geographical location and distribution of metropolitans and primary roads, and county-level factors such as median household income could affect the rate as well. The drilled counties might be relatively less developed at the same time, so the development focusing and safety focusing policies could be implemented respectively.

There are also several limitations in this study. The most important one could be the availability of outcome variable. The ideal outcome for the analysis should be crashes of drilling-related heavy trucks, which was not available through public source. Using total crashes as outcome variable might mixed up all the crashes that were not related to drilling activities. Second, this ecological study suffered from potential ecological fallacy. The county-level data may not truly reflect the distribution of sociodemographic data, and differences of these factors might still exist within the range of a county. Another limitation was the selection of covariates. There were only
limited previous studies and very few of them considered the impact of other covariates. Therefore, the selection of covariates was referred to other epidemiological studies with different outcomes\textsuperscript{12}. This limitation led to some unexpected results of association in the study, while residual confounding might still exist and need further exploration. In addition, the variables in the study did not contain enough geographical information. Failing to obtain the location of crash incident might mix all crashes together within a county, without considering the distance from well construction site, and might partially reduce the validity of the inferred association. The study provided the evidence of significant impact from primary roads distribution, so it is natural to suspect the existence of other unmeasured spatial confounders. More future works can be done in order to promote the validity of the study. More detailed traffic data including the cause of crashes, whether heavy truck related, the road condition, etc. could better reveal the true association. Besides, the location information of crashes and well sites could be collected to conduct geographical analysis investigating the relationship between traffic accident density and drilling activity intensity\textsuperscript{21}. The matching analysis could also be improved by considering matching on more sociodemographic and spatial factors with potential impacts.

When the oil and gas exploration industry was still expanding, the raised public health concerns should not be ignored\textsuperscript{3}. Although this study found a slight possible “protecting effect” of drilling intensity of unconventional wells, it was not suggesting the effect on
traffic could be underrated. Evidence showed that the oil and gas industry may associated with a high rate of fatal injuries, and highway crashes and work-related motor vehicle-related contributed significant proportion\textsuperscript{13, 20-22}. The traffic accidents were not the risk only related to the industry-related workers and drivers, but also could be a concern of the communities where oil and gas wells were located\textsuperscript{15}. Further studies are necessary to provide more knowledge about public health concerns related to unconventional oil and gas development. This study suggests that the traffic department, including department of public safety could strengthen the supervision and make more detailed statistics for deeper analysis. This study also suggests that the oil and gas industry should adopt alternative transportation methods to reduce the fatality of drivers and passengers via neighborhood of drilling sites instead of focusing on developing regardless of public health concerns.
Tables

Table 1. Distribution of sociodemographic factors among ever drilled and not drilled counties in Ohio in predrilling period (2004–2009).

<table>
<thead>
<tr>
<th>Sociodemographic factors</th>
<th>Ever drilled counties (^a)</th>
<th>Not drilled counties</th>
<th>P-value (^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=24</td>
<td>N=64</td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td>95228 (84587)</td>
<td>143825 (241124)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Percent white</td>
<td>94.55 (3.79)</td>
<td>92.39 (6.81)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Percent black</td>
<td>3.30 (3.55)</td>
<td>4.25 (6.00)</td>
<td>0.026</td>
</tr>
<tr>
<td>Percent Hispanic</td>
<td>1.01 (0.75)</td>
<td>2.20 (1.97)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Percent males</td>
<td>49.38 (1.93)</td>
<td>49.36 (1.26)</td>
<td>0.912</td>
</tr>
<tr>
<td>Percent with health insurance</td>
<td>86.00 (6.79)</td>
<td>88.64 (2.45)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Percent of population aged 15-29 years old</td>
<td>18.97 (2.57)</td>
<td>19.88 (3.48)</td>
<td>0.001</td>
</tr>
<tr>
<td>Percent of population had high school degree</td>
<td>83.69 (6.86)</td>
<td>84.94 (4.38)</td>
<td>0.044</td>
</tr>
<tr>
<td>Percent of population had bachelor’s degree</td>
<td>15.19 (6.10)</td>
<td>17.15 (7.61)</td>
<td>0.003</td>
</tr>
<tr>
<td>Median household income (in US dollars)</td>
<td>42341 (8452)</td>
<td>45934 (9107)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Note:

a. Table values are mean (standard deviation) for variables.

b. P-value is for t-test of difference between ever drilled and not drilled counties.

Table 2. Distribution of sociodemographic factors among ever drilled and not drilled counties in Ohio in predrilling period (2010–2017).

<table>
<thead>
<tr>
<th>Sociodemographic factors</th>
<th>Ever drilled counties (^a)</th>
<th>Not drilled counties</th>
<th>P-value (^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=24</td>
<td>N=64</td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td>94330 (83350)</td>
<td>145728 (244568)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Percent white</td>
<td>94.23 (3.94)</td>
<td>91.74 (7.51)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Percent black</td>
<td>3.22 (3.43)</td>
<td>4.48 (6.34)</td>
<td>0.001</td>
</tr>
<tr>
<td>Percent Hispanic</td>
<td>1.43 (0.99)</td>
<td>2.75 (2.24)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Parameter</td>
<td>Rate ratios estimate</td>
<td>95% CI</td>
<td>P-value</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>----------------------</td>
<td>------------------</td>
<td>---------</td>
</tr>
<tr>
<td><strong>Continuous model</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline rate</td>
<td>0.0254</td>
<td>(0.0254, 0.0254)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Total horizontal wells a</td>
<td>0.9919</td>
<td>(0.9915, 0.9924)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Binary model</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline rate</td>
<td>0.0258</td>
<td>(0.0258, 0.0259)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Ever drilled (yes vs no)</td>
<td>0.8966</td>
<td>(0.8936, 0.8996)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Categorical model</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline rate</td>
<td>0.0257</td>
<td>(0.0257, 0.0258)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Intensity (medium vs low)</td>
<td>0.9114</td>
<td>(0.9076, 0.9151)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Intensity (high vs low)</td>
<td>0.8514</td>
<td>(0.8449, 0.8578)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Note:

a. Unit for total horizontal wells was per 10 wells.

Table 3. Rate ratios estimate from crude regression models for association between drilling activity and traffic incidents in Ohio 2010-2017 (n=704 county-years)

a. Table values are mean (standard deviation) for variables.
b. P-value is for t-test of difference between ever drilled and not drilled counties.
Table 4. Rate ratios estimate from multivariable regression models for association between drilling activity and traffic incidents in Ohio 2010-2017 (n=704 county-years)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Rate ratios estimate (95% Confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Continuous model</td>
</tr>
<tr>
<td>Baseline rate</td>
<td>0.0061 (0.0056, 0.0067)</td>
</tr>
<tr>
<td>Well number variable</td>
<td>0.9944 (0.9939, 0.9949)</td>
</tr>
<tr>
<td>Road across</td>
<td>1.1270 (1.1220, 1.1322)</td>
</tr>
<tr>
<td>% white</td>
<td>0.9989 (0.9986, 0.9993)</td>
</tr>
<tr>
<td>% Hispanic</td>
<td>0.9922 (0.9914, 0.9930)</td>
</tr>
<tr>
<td>% males</td>
<td>1.0197 (1.0180, 1.0214)</td>
</tr>
<tr>
<td>% health insurance</td>
<td>1.0092 (1.0086, 1.0097)</td>
</tr>
<tr>
<td>% population 15-29 years</td>
<td>0.9996 (0.9989, 0.0002)</td>
</tr>
<tr>
<td>% bachelor’s degree</td>
<td>1.0101 (1.0094, 1.0107)</td>
</tr>
<tr>
<td>Median household income</td>
<td>0.8891 (0.8851, 0.8933)</td>
</tr>
</tbody>
</table>

Note:

a. Well number variable refer to total horizontal wells in this model, unit: per 10 wells.
b. Well number variable refer to ever drilled (yes vs no) in this model.
c. Well number variable refer to medium vs low drilling intensity in this grid.
d. Well number variable refer to high vs low drilling intensity in this grid.
e. Unit for median household income was per 10,000 US dollars.

Table 5. Rates ratios estimate from multivariable regression model by different incident types in Ohio 2010-2017 (n=704 county-years)

<table>
<thead>
<tr>
<th>Incident type</th>
<th>Rates ratio estimate (95% Confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Continuous model</td>
</tr>
<tr>
<td>Total crashes</td>
<td></td>
</tr>
<tr>
<td>Well number variable</td>
<td>0.9944 (0.9939, 0.9949)</td>
</tr>
<tr>
<td>Road across</td>
<td>1.1270 (1.1220, 1.1322)</td>
</tr>
</tbody>
</table>
### Fatal crashes

- **Well number variable**
  - 0.9985 (0.9927, 1.0043) 0.9323 (0.8763, 0.9919) 0.9106 (0.8477, 0.9782)

- **Road across**
  - 0.9257 (0.8684, 0.9869) 0.9438 (0.8834, 1.0084) 0.9433 (0.8832, 1.0074)

### Injury crashes

- **Well number variable**
  - 0.9965 (0.9956, 0.9974) 0.9770 (0.9692, 0.9849) 0.9748 (0.9660, 0.9837)

- **Road across**
  - 1.1196 (1.1095, 1.1299) 1.1273 (1.1167, 1.1378) 1.1240 (1.1135, 1.1345)

### PDO crashes

- **Well number variable**
  - 0.9937 (0.9931, 0.9942) 0.9092 (0.9050, 0.9135) 0.9398 (0.9348, 0.9448)

- **Road across**
  - 1.1308 (1.1248, 1.1367) 1.1599 (1.1535, 1.1661) 1.1415 (1.1353, 1.1476)

**Note:**

- a. Well number variable refer to total horizontal wells in this model, unit: per 10 wells.
- b. Well number variable refer to ever drilled (yes vs no) in this model.
- c. Well number variable refer to medium vs low drilling intensity in the upper grid for each crash type.
- d. Well number variable refer to high vs low drilling intensity in lower grid for each crash type.

#### Table 6. Measurement of goodness of matching

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Crash rate</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD) for drilled counties a</td>
<td>0.0231 (0.0050)</td>
<td>95050 (85698)</td>
</tr>
<tr>
<td>Mean (SD) for not drilled counties a</td>
<td>0.0230 (0.0049)</td>
<td>93954 (80012)</td>
</tr>
<tr>
<td>P-value b</td>
<td>0.9245</td>
<td>0.9637</td>
</tr>
<tr>
<td>Standardized difference before matching</td>
<td>-0.3922</td>
<td>-0.2719</td>
</tr>
<tr>
<td>Standardized difference after matching</td>
<td>0.0286</td>
<td>0.0060</td>
</tr>
<tr>
<td>Variance ratio before matching</td>
<td>1.2205</td>
<td>0.1249</td>
</tr>
<tr>
<td>Variance ratio after matching</td>
<td>1.0406</td>
<td>1.1472</td>
</tr>
</tbody>
</table>

**Note:**

- a. The statistics are for the observations after matching.
- b. P-value for t-test of drilled and not drilled counties after matching.
### Table 7. Rate ratios estimate from multivariable regression models among matched counties in Ohio 2010-2017 (n=384 county-years)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Rate ratios estimate (95% Confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Continuous model</td>
</tr>
<tr>
<td>Baseline rate</td>
<td>0.0206 (0.0178, 0.0238)</td>
</tr>
<tr>
<td>Well number variable</td>
<td>0.9976 (0.9971, 0.9981) a</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Road across</td>
<td>1.1787 (1.1712, 1.1864)</td>
</tr>
<tr>
<td>% white</td>
<td>1.0026 (1.0019, 1.0034)</td>
</tr>
<tr>
<td>% Hispanic</td>
<td>1.0012 (0.9990, 1.0034)</td>
</tr>
<tr>
<td>% males</td>
<td>0.9873 (0.9848, 0.9899)</td>
</tr>
<tr>
<td>% health insurance</td>
<td>1.0060 (1.0053, 1.0067)</td>
</tr>
<tr>
<td>% population 15-29 years</td>
<td>1.0070 (1.0061, 1.0080)</td>
</tr>
<tr>
<td>% bachelor’s degree</td>
<td>1.0025 (1.0016, 1.0034)</td>
</tr>
<tr>
<td>Median household income</td>
<td>0.9324 (0.9265, 0.9384)</td>
</tr>
</tbody>
</table>

Note:

a. Well number variable refer to total horizontal wells in this model, unit: per 10 wells.

b. Well number variable refer to ever drilled (yes vs no) in this model.

c. Well number variable refer to medium vs low drilling intensity in this grid.

d. Well number variable refer to high vs low drilling intensity in this grid.

e. Unit for median household income was per 10,000 US dollars.

### Table 8. Rate ratios estimate from multivariable regression models using continuous road variable in Ohio 2010-2017 (n=704 county-years)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Rate ratios estimate (95% Confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Continuous model</td>
</tr>
<tr>
<td>Baseline rate</td>
<td>0.0061 (0.0056, 0.0067)</td>
</tr>
<tr>
<td>Well number variable</td>
<td>0.9928 (0.9923, 0.9933) a</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Width</td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------</td>
</tr>
<tr>
<td><strong>Distance</strong></td>
<td>0.9736 (0.9727, 0.9745)</td>
</tr>
<tr>
<td><strong>% white</strong></td>
<td>0.9998 (0.9994, 1.0001)</td>
</tr>
<tr>
<td><strong>% Hispanic</strong></td>
<td>0.9905 (0.9897, 0.9913)</td>
</tr>
<tr>
<td><strong>% males</strong></td>
<td>1.0202 (1.0185, 1.0219)</td>
</tr>
<tr>
<td><strong>% health insurance</strong></td>
<td>1.0102 (1.0096, 1.0107)</td>
</tr>
<tr>
<td><strong>% population 15-29 years</strong></td>
<td>0.9989 (0.9982, 0.9996)</td>
</tr>
<tr>
<td><strong>% bachelor’s degree</strong></td>
<td>1.0112 (1.0106, 1.0118)</td>
</tr>
<tr>
<td><strong>Median household income</strong></td>
<td>0.8784 (0.8742, 0.8824)</td>
</tr>
</tbody>
</table>

**Note:**

a. Well number variable refer to total horizontal wells in this model, unit: per 10 wells.

b. Well number variable refer to ever drilled (yes vs no) in this model.

c. Well number variable refer to medium vs low drilling intensity in this grid.

d. Well number variable refer to high vs low drilling intensity in this grid.

e. Unit for distance was per 10 km.

f. Unit for median household income was per 10,000 US dollars.
Figure 1. Total horizontal well numbers by Ohio counties
Figure 2. Mean incident rates across counties and time by drilling activity

(A) Time trend for total crash rate. (B) Time trend for fatal crash rate. (C) Time trend for injury crash rate. (D) Time trend for PDO crash rate.
Figure 3. Total incident rates of 2010 to 2017 by Ohio counties
Figure 4. Result of Spearman correlation analysis of demographic variables
Figure 5. Result of matched counties in Ohio
Reference


22. Additional Road Investments Needed to Support Oil and Gas Production and Distribution in North Dakota; Upper Great Plains Transportation Institute, North Dakota State University: 2010.


30. 2020 TIGER/Line Shapefiles (machinereadable data files) / prepared by the U.S. Census


