Fracking Boom and Respiratory Health: Evidence from Texas *

Ziyue Xu[1]

[1] *Department of Economics, University of Texas at Austin, Austin, TX, USA*

August 30, 2023

Abstract

This paper investigates the impact of fracking (shale oil and gas) on respiratory health conditions, using data from Texas and a continuous difference-in-difference design. I find that compared to a county with the 25th percentile of the reserves and comparing post to pre fracking periods, a county with the 75th percentile of the reserves has 0.76 (5.31) more asthma (general respiratory disease) hospitalizations per 100,000 residents, which is a 0.82% (0.55%) increase compared to the average. These impacts are the largest for children and teenage. By back-of-envelop calculation, the fracking boom brought approximately 15,000 extra asthma hospitalizations and 103,000 extra respiratory disease hospitalizations to fracking counties in Texas between 2005 and 2014. I explore three channels to explain the mechanisms of these findings: migration, air pollution, and water pollution. Air and water pollution channels are consistent with the health impact, while migration is not.

Keywords: Fracking, respiratory health, air pollution, water pollution, difference-in-difference

*Corresponding author: `ziyue353@utexas.edu`. Address: 2225 Speedway (Office 2.124), Austin, TX 78705. I am grateful to Manuela Angelucci, Mike Geruso, and Marika Cabral for their invaluable guidance and support. For helpful comments, I thank Richard Murphy, Nir Eilam, Bokyung Kim, Alina Kovalenko, Jinyeong Son, Prankur Gupta and all seminar participants at the University of Texas at Austin. I would also like to thank the Texas Health Care Information Collection Center and DrillingInfo for generously providing the data. All errors are my own. The Conclusions of this research do not necessarily reflect the opinion or official position of the Texas Health Care Information Collection Center, Texas Department of State Health Services, or the State of Texas.

1 Introduction

Although the fracking boom provides almost a million jobs in the U.S., it also has externalities on human health and the environment. [1] When fracking, water, and other chemicals are forced into shale rock to fracture it and allow the gas and oil to be tapped, which has a high risk of polluting the groundwater. [2] Besides, drilling by fracking and the process of oil and natural gas production and transportation may also increase the air pollutant concentration and PM level in the vicinity of an active unconventional well. There have already been solid environmental science and economic research showing evidence that fracking indeed pollutes air and water (U.S. EPA, 2016; Kim et al., 2016; Butkovskyi et al., 2017; Hill and Ma, 2022; Zhang et al., 2020). Since about 17.6 million Americans live within one mile of an active drilling site Czolowski et al. (2017), there is a clear need to ascertain the effects of the fracking boom on human health in every aspect based on the environmental impacts it brings to the communities.

Prior to the development of fracking technology, shale oil, and gas were considered essentially useless because the cost to extract them was prohibitively expensive. Beginning in the early 2000s, the advancements in fracking technology and horizontal drilling, along with increased energy prices, motivated the shale oil and gas boom in the United States. The number of unconventional drilling sites increased from around 26,000 to more than 350,000 between 2000 and 2019. The crude oil production in Texas triples in the same period.

This project uses inpatient hospitalization data in Texas and a continuous difference-in-difference research design to detect the causal impacts of the fracking boom on residents' respiratory health status. Compared to previous literature discussing fracking and human health, this project uses predicted oil and natural gas reserves to measure a community's fracking potential instead of using other direct measures like the number of wells and oil production. Using drilling as a direct measure of exposure to the fracking boom has omitted variable issues. For example, companies' decisions to extract oil and gas may be endogenous, and counties with disadvantaged economic status may be more willing to accept fracking. The social-economic characteristics that motivate counties to

[1] Hydraulic fracturing, commonly referred to as "fracking," is a process that uses high-pressure fluid injections to shatter rock formations and extract oil and natural gas from the shale layer.

[2] Shale is a porous and sedimentary rock that sits miles beneath the ground and holds enormous quantities of oil and natural gas.

take fracking may also correlate to human health outcomes. By using predicted oil and natural gas reserves assigned by nature before fracking as a technology is developed, the omitted variable issue discussed above can be dealt with. To construct the reserve measure at the commuting zone level, I use basin-level raw data of predicted shale oil reserve and natural gas reserve (separately) from the U.S. Energy Information Administration (EIA). Then I assign these reserves to each commuting zone by the size of the overlapped area between the commuting zones and shale plays. Also, little literature finds robust evidence of the fracking boom on adults. Most research focuses on detecting the health impacts of fracking on vulnerable age groups, especially babies and pregnant women. To my understanding, this is the first project to provide robust causal evidence of the health impacts of the fracking boom among different age groups, including young adults.

This project finds that compared to a county with the 25th percentile of the reserves and comparing post to pre fracking periods, a county with the 75th percentile of the reserves has 0.76 (5.31) more asthma (general respiratory disease) hospitalizations per 100,000 residents, which is a 0.82% (0.55%) increase compared to the average. These negative impacts are also robust by samples of different age groups, but the magnitude varies. These impacts are much more prominent if restricting the sample population of residents between 0 and 17 years old than in other age groups.

This project also explores the mechanisms of these findings through three different channels. First, I use American Community Survey (ACS) data to examine whether fracking influences migration patterns. Since there are no significantly more non-adult inflow migrants to counties with higher reserves, migration should not be the leading cause of these impacts. Instead, air and water pollution might be the potential mechanism to explain these findings. Using the NASA Terra satellite aerosol optical depth (AOD) data, fracking drilling increases the AOD level within 2km of an unconventional well by 0.727% compared to the average. This number is smaller than the 2.19% increased AOD level found in Zhang et al. (2020), another paper that studies the air pollution impacts of fracking by using Pennsylvania data. To get this estimate, I pull AOD data within circle areas with 20km as the radius and each well as the center of the circle. I use locations in the ring 10-20km away from each well as the control group. For robustness, I also try different treatment groups, locations within 0-2km ring, 0-5km ring, and 0-10km ring. I also add the wind direction as the weight of the regression. The wind direction weight is measured by the cosine function of the angle between two vectors. One is the vector from the well to the AOD reporting location, and another vector is the wind direction from the well. Besides, I also find that fracking increases the concentration of drilling-related chemical substances in the purchased groundwater by using chemical substance sample reports from Texas Drinking Water Watch System. However, it is difficult to calculate the proportions of the contribution to these health impacts between the air and water pollution channels. By back-of-envelop calculation, the fracking boom brings approximately 15,000 extra asthma hospitalizations and 103,000 respiratory disease hospitalizations to fracking counties in Texas during the period between 2005 and 2014.

The rest of the paper is organized as follows. Section 2 presents a review of the literature discussing fracking and health. Section 3 provides background information on the fracking boom. Section 4 and 5 describe the data and research design. Section 6 presents main results and robustness checks. Section 7 explores potential mechanisms, and Section 8 concludes.

2 Literature Review

There are currently two categories of literature investigating fracking and health. The first discusses fracking and infant health. There are at least three reasons to focus mainly on infant health in probing the health effects of exposure to hydraulic fracturing(Currie et al., 2017). First, there is growing evidence that the fetus is vulnerable to a range of maternal pollution exposures (Mattison et al., 2003; Chay and Greenstone, 2003; Currie et al., 2009). Second, the fetus is in utero for at most nine months, which makes it possible to pinpoint the timing of potential exposure. Third, birth data are usually provided with precise information, including birth weight, mother's residential location, and other health outcomes. According to Pennsylvania infant health data studies, in utero exposure to fracking sites has been found to have adverse effects on infant weights (Currie et al., 2017; Denham et al., 2019). A positive association between fracking and early infant mortality is also found in the literature (Busby and Mangano, 2017).

Another category of literature addresses fracking and human diseases. A focus of these diseases is asthma. An association between asthma, especially pediatric asthma, is found in the current medical and public health literature (Rasmussen et al., 2016; Willis et al., 2020). Other literature finds an association between fracking and pneumonia, but the results are only robust among the elderly (Peng et al., 2018). Fracking counties are also found to have more cardiology and neurology hospitalizations in medical literature (Jemielita et al., 2015). The current literature discussing fracking and diseases has the following limitation. Almost all these papers use the number of fracking wells or oil and natural gas production as the direct measure of exposure to the fracking boom, which might bring omitted variable issues. The decision to extract oil and gas by companies may be endogenous. Since fracking has been a controversial topic in the United States for years, struggling counties may be more willing to accept fracking to boost their economy. These characteristics of counties that motivate to take fracking may also correlate to residents' health status.

This project contributes to the literature in the following three ways. First, I use predicted oil and gas reserves as a continuous measure of the fracking boom to deal with the omitted variable issues existing in previous literature. Second, this project's findings expand the health impacts of the fracking boom on pediatric asthma to general respiratory tract diseases compared to Willis et al. (2020). Third, to my understanding, this is the first project which finds robust impacts of fracking on respiratory health among all age groups instead of only vulnerable age groups.

3 Background

3.1 Fracking Boom in Texas

Hydraulic fracturing commonly referred to as "fracking", is a process that uses high-pressure fluid injections to shatter rock formations and extract oil and natural gas. Beginning in the early 2000s, the advancements in fracking technology and horizontal drilling and increased energy prices motivated the shale oil and gas boom in the United States.

Shale is a type of porous and sedimentary rock that sits miles beneath the ground and holds enormous quantities of oil and natural gas. Unlike conventional deposits, which are found in pockets, shale oil and gas are dispersed throughout the formation in thin layers. Therefore conventional vertical drilling is typically not considered a feasible method of extracting resources from shales (Kovalenko, 2023). Fracking technology has been around since the 1940s, but it wasn't until it was combined with horizontal drilling that shale oil and gas became accessible. By drilling long horizontal wells, more production can be delivered from a single well and minimize the surface footprint. These technological innovations, together with high oil and gas prices in the early 2000s, motivated localized fracking booms in many states with reserves across the united states.

Texas plays a significant role in the U.S. energy market, and this state sits on top of four major shale basins: Permian, Barnett, Eagle Ford, and Haynesville-Bossier. Figure 1 presents the shapes and locations of the four shale basins. Barnett and Eagle Ford are within the boundary of Texas among these four basins. At the same time, Permian is a basin crossing both Texas and New Mexico, and Haynesville-Bossier is a basin crossing both Texas and Louisiana. By introducing the new extraction technology, "fracking", these shale oil and gas reserves became extremely valuable, which motivated Texas counties on top of the shale to actively apply fracking in drilling. In this paper, I use 2005 as the year Texas counties started actively applying fracking technology in drilling by following Kovalenko (2023). As shown in Figure 3, the number of unconventional wells (fracking wells) located on top of the shale increased by more than 1000% from 2004 to 2014 (Panel A). However, unconventional wells remained stable from 1994 to 2019 (Panel B). This figure implies that the introduction of fracking doesn't influence the traditional drilling activities too much but booms the fracking activities on top of the shale plays. As a result of the increasing fracking drilling activities, Texas saw a drastic increase in crude oil and natural gas production, as shown in Figure 4. Especially for crude oil, the production tripled between 2005 and 2014.

3.2 Measuring Exposure to the Boom

One of the main challenges in estimating the causal impacts of the fracking boom on human health outcomes with county-level data is that the decision to extract oil and gas by companies (or getting permission for fracking drilling by local communities) may be endogenous. Struggling communities may be more willing to accept fracking to boost employment or tax revenue than rich communities. Therefore, using the number of wells or total oil and natural gas production in a county to measure the exposure to the fracking boom might introduce omitted variable bias if the same characteristics that ultimately led to the extraction of oil and gas reserves also affect residents' health outcomes.

I instead use predicted shale oil and gas reserves per capita as a measure of the fracking potential of an area by following the approach in three labor literature (Kovalenko, 2023; Cascio and Narayan, 2015; Michaels, 2011). By following Kovalenko (2023), I define a local area as a commuting zone since commuting zones are geographic units of analysis intended to reflect the local economy where people live and work more closely. Especially for fracking areas in Texas, several counties form one commuting zone because these areas are less populated and may share some medical resources. To construct the reserve measure at the commuting zone level, I use basin-level raw data of predicted shale oil reserve and natural gas reserve (separately) from the EIA. Then I assign these reserves to each commuting zone by the following steps. First, I overlay shale maps with Texas commuting zone boundary shapefiles, which are shown in Figure 5. Second, I allocate oil and gas reserves to commuting zones based on the share of each shale that this commuting zone represents. Third, I convert oil and natural gas reserves into one standard metric defined by millions of British Thermal Units (MMBTUs). Finally, I divide the variable in MMBTU by the baseline population of each commuting zone in 2000 to make the variable per capita.

Figure 6 displays geographic variation in the fracking potential across commuting zones in Texas, measured by the predicted shale oil and gas reserve per capita. The reserves per capita of commuting zones vary from 0 to 782 million MMBTU. The darker red a commuting zone shows in this figure, the higher reserve per capita and the higher the fracking potential it has. There are clusters of extremely high-reserve areas in the west and south, where Permian Basin and Eagle Ford Basin are located. There are two other clusters of high-reserve regions found in the middle and east, where Barnett Basin and Haynesville-Bossier Basin locates. My research design relies on the assumption that reserves are a good proxy for fracking drilling activity and, ultimately, the extent of the negative health impacts on residents. To support this assumption, I regress newly drilled fracking wells on interactions between reserve per capita and year indicators and year. Year and commuting zone fixed effects are included in this regression. The coefficients of this regression are plotted in Figure 7. Commuting zones with high reserves saw significant increases in fracking wells, suggesting that underlying reserves are a good proxy for subsequent drilling activities. The number of fracking wells increased significantly around 2005 and peaked between 2012 and 2014. Before 2005, the coefficients are precisely estimated as zeros. This figure also shows that the starting time of the fracking boom in Texas is 2005, when counties with high reserves start actively drilling wells by fracking, which is consistent with Kovalenko (2023). Although different counties might start actively taking fracking in drilling at different years [3], I use 2005 as a general starting point of the fracking boom in Texas in my principal analysis.

4 Data

The main data used in this project comes from Texas Inpatient Public Use Data Files (PUDF). This is quarterly reported data of hospitalization records based on inpatient discharges collected from about 700 hospitals since 1999, covering hospitals in each Texas county. Each piece of record in this data includes the patient's age, race, gender, residential county and ZIP code, length of stay, charges, and diagnosis details. My main sample uses hospitalization records correlated to respiratory health diseases between 2000 to 2014. I didn't include data after 2015 in my analysis because PUDF started using ICD-10-CM [4] coding for categorizing diseases in 2015. Before that, PUDF used ICD-9-CM as the classification standard of diseases. Since I generate samples of hospitalizations of respiratory diseases in my analysis, I give up the data after 2014 to keep my definition of the group of respiratory diseases consistent over the years. To protect the privacy of inpatients, PUDF hides the last two digits if a ZIP code has fewer than 30 discharges. PUDF also hides the whole five digits if a hospital has fewer than 50 discharges or if a hospital has fewer than five discharges of a particular gender. Therefore, I only include the residential county of the inpatients in my analysis.

The energy and drilling data used in this project come from two sources. The first one is Enverus DrillingInfo Database [5], which is data provided by a private company that focuses on energy sector analysis. This database tracks each active well in the U.S., including unconventional (fracking) and conventional wells. Plentiful information about each well is provided in this data,

[3] I also estimate a staggered difference-in-difference equation to check the robustness of the main results as discussed below.

[4] World Health Organization (WHO) authorized the publication of the International Classification of DiseasesExternal 10th Revision (ICD-10), which was implemented for mortality coding and classification from death certificates in the U.S. in 1999. The U.S. developed a Clinical Modification (ICD-10-CM) for medical diagnoses based on WHO's ICD-10, and CMS used a new Procedure Coding System (ICD-10-PCS) for inpatient procedures. ICD-10-CM replaces ICD-9-CM, volumes 1 and 2, and ICD-10-PCS replaces ICD-9-CM, volume 3.

[5] I obtained access to the data through a special agreement with DrillingInfo

including location, rig dates, drilling type, and production. The second source is the U.S. Energy Information Administration (EIA). My main independent variable, predicted oil and natural gas reserve, is generated by raw files from EIA websites, including the predicted shale oil and natural gas reserve of each main basin in the U.S. and the shapefiles of the shale plays.

I also use data from American Community Survey (one-year estimates) to get demographic control variables and explore the potential mechanism of migration. I also collected chemical substance sample report data from the Texas Drinking Water Watch System to identify the impacts of the fracking boom on drinking water quality. I use NASA satellite products to identify the fracking boom's impacts on air pollution. More details of how these data are used will be discussed in Section 7.

5 Research Design

5.1 Standard Difference-in-Difference

My main specification is a standard continuous difference-in-difference model. I compare changes in hospitalizations across local residents with high exposure to the fracking boom (high reserve) to changes in outcomes across local residents with lower exposure (low reserve). According to the event study pattern shown in Figure 7 and existing literature studying the Texas fracking boom, I use 2005 as the cutoff year when Texas started actively applying fracking in drilling wells. I estimate the following equation:

$$y_{ijt} = \sum_{t \neq 2004} \beta_k \cdot I_{k=t} \cdot Reserve_i + \theta X_{it} + \gamma_i + \delta_t + \epsilon_{ijt} \tag{1}$$

In the above equation, y_{ijt} is the number of hospitalizations per 100,000 residents related to either asthma or respiratory diseases in commuting zone i, county j, and year t. $Reserve_i$ is a continuous normalized measure of predicted shale oil and natural gas reserves per capita in commuting zone i. The term $I_{k=t}$ represents a set of dummy variables equal to 1 in year t. X_{it} are commuting zone covariates in year t, including the log value of the median household income and poverty rate [6]. γ_i and δ_t are commuting zone fixed and time fixed effects. The standard errors are clustered by commuting zone level. I normalize β_{2004} to zero, so all coefficients can be interpreted as changes relative to the inpatients hospitalized in 2004, which is the last never-treated cohort. Inpatients after 2005 are partially treated strictly speaking because different counties actively take fracking in different years. Therefore, I also run regressions by a staggered difference-in-difference framework later. However, following the standard difference-in-difference framework, I consider the inpatient cohorts after 2004 as generally treated groups.

Besides equation (1), I also estimate another equation to get the point estimate of the association between reserve and hospitalizations, where $Treat_t$ equals one if a county was in 2005 or later.:

$$y_{ijt} = \beta_0 + \beta_1(Reserve_i \times Treat_t) + \theta X_{it} + \gamma_i + \delta_t + \epsilon_{ijt} \tag{2}$$

5.2 Staggered Difference-in-Difference

Although there has been a significant increase in the number of fracking wells in Texas since 2005, according to Figure 7, different Texas counties actively take fracking in different years. As in Figure 8, I plot the numbers of yearly fractured wells in the four counties with the most fracking wells in Texas. Tarrant county started fracking in 2002, earlier than the general starting point of the Texas fracking boom in 2005. However, Karnes county and Reeves county actively applied fracking in drilling in 2009, and Dimmit county started in 2010. To more precisely estimate the impacts of the fracking boom and also check the robustness among different specifications of identification strategies, I also estimate a staggered difference-in-difference equation as follows:

$$y_{ijt} = \sum_{T_i \neq 0} \beta_k \cdot I_{k=T_i} \cdot Reserve_i + \theta X_{it} + \gamma_i + \delta_t + \epsilon_{ijt} \tag{3}$$

In the above equation, the only difference from the standard difference-in-difference framework (equation (1)) is the term $I_{k=T_i}$. T_i represents the normalized year for county i, which equals the year of that inpatient cohort minus the fracking boom starting year of county i plus one. For example, if the fracking boom starting year of county i is 2009, then T_i = t-2009+1. I define the

[6] These control variables come from the American Community Survey (1-year estimate).

starting year of the fracking boom of a county as the year when this county has more than ten newly drilled fracking wells. [7]

Besides equation (3), I also estimate another equation as follows to get the point estimate that shows the certain change of reserve is equivalent to how much change of the number of hospitalizations, where $Treat_{it}$ equals one if a county was in or after its fracking starting year.

$$y_{ijt} = \beta_0 + \beta_1(Reserve_i \times Treat_{it}) + \theta X_{it} + \gamma_i + \delta_t + \epsilon_{ijt} \tag{4}$$

The staggered difference-in-difference framework also provides me with a way to deal with the data limitation in the section of exploring mechanisms of my main findings, details of which will be discussed later in this project [8].

6 Results

6.1 Results of the Difference-in-Difference Design

The coefficients of the standard continuous difference-in-difference framework (equation (1)) are presented in Figure 9. Panel A shows that compared to counties with lower reserves, counties with higher reserves have significantly more asthma hospitalizations per capita since 2006 (one year after the cutoff). Panel B shows that compared to counties with lower reserves, counties with higher reserves have significantly more respiratory disease hospitalizations per capita since 2007 (two years after the cutoff). Except for the year 2000 in Panel A, the coefficients of the pre-trend period between 2000 and 2004 are not statistically significant from zero, which implies that the assumption of this continuous difference-in-difference design, counties with higher reserve and counties with lower reserve have parallel patterns of respiratory disease hospitalizations, is generally valid. I also do heterogeneous analysis by dividing the sample of the whole population into four

[7] I don't use the year when a commuting zone has the first fracking well as the starting year of the boom because, for some commuting zones, they started fracturing one or two wells in a year and then stopped keeping doing it until a couple of years later. So having ten wells is a better measure of actively applying fracking in the drilling of an area.

[8] A couple of data sets I used for detecting potential migration pattern and water pollution pattern are only available since 2004 or 2005. If I still apply a standard difference-in-difference strategy, I have zero data points in the pre-trend.

different age groups and replicating the regression of equation (1), results of which are presented in Figure 10 and 11. Figure 10 plots the effects of fracking boom on asthma hospitalizations by age groups. All four age groups show statistically significantly more asthma hospitalizations in counties with higher reserve. The four panels in 10 imply that the negative impacts found of fracking on asthma are robust among different age groups, but the magnitude of the impacts varies. The fracking boom has much stronger impacts on non-adults and people at retirement age than those on adults from 18-64 years old. Figure 11 plots the effects of fracking boom on general respiratory disease hospitalizations by age groups and shows almost the same patterns as Figure 10.

I also estimate equation (2) to get one estimate for all the analyses above, the results of which are shown in Table 1. Panel A presents all the results of asthma hospitalizations, and Panel B presents those of respiratory disease hospitalizations. The first and second column of Table 1 shows the regression results of equation (2) with the sample of the whole population. Including control variables, poverty rate, and log median household income almost doesn't influence the number and significance of the coefficients. A unit increase of reserve is equivalent to 0.017 more asthma hospitalization and 0.118 more general respiratory disease hospitalization. To transfer these coefficients into a number easily to be understood, I compare the county with 25 percentile of reserve (1.282 MMBTU) and the county with 75 percentile of reserve (46.817 MMBTU). Therefore, there exist approximately 45 units of reserve gap, and based on these coefficients, 45 units more reserve is equivalent to 0.76 (5.31) more asthma (respiratory disease) hospitalizations per 100,000 residents, which is a 0.82% (0.55%) increase compared to the average.

To prove that the increasing patterns of hospitalizations are specific to respiratory tract diseases, I also replicate the regressions of equation (1) by replacing the Y variable of respiratory disease hospitalizations with the total number of hospitalizations. The results are plotted in Figure 14, which doesn't show any evidence that counties with higher reserve have more total number of hospitalizations. I also replicate equation (1) by restricting the sample to several categories of non-respiratory tract disease hospitalizations. In Figure 15, I plot the event study results of four types of non-respiratory diseases, mental disorders, endocrine, diseases of the circulatory system, and diseases of the digestive system, and I find no significant patterns in non-respiratory disease hospitalizations in Texas.

The results of the staggered difference-in-difference framework (equation (3)) are plotted in Figure 12, and I find that both asthma hospitalizations and respiratory disease hospitalizations in counties with higher reserve increase right after the cutoff year. These results imply that these health impacts are robust among different identification specifications.

6.2 Robustness Checks

To check the robustness of the results of the main identification strategy (equation (1) and (2)), I also do a few robustness checks. The main idea is to redefine the sample included in the analysis. First, I remove the counties with at least one oil refinery in my sample. Oil refineries are a big source of air pollution in Texas. Currently, there are 26 active oil refineries in Texas, which are located in 13 different counties. Some counties have more than one oil refinery. For example, Harris county has five ones. Negative impacts of oil refineries have been shown in literature, including air pollution, water pollution, and heavy metals in the soil (Logsdon, 1985; Jaramillo and Muller, 2016; Croquer et al., 2016). Especially, there exists medical literature that has already found that children living in the polluted petrochemical area show an increase in wheezing symptoms, a decrease in lung function, and an increase in bronchial inflammation (Rusconi et al., 2011). Therefore, including those counties with at least one oil refinery in the sample might result in a downward bias of the estimates because most oil refineries are set in non-fracking counties. To check for this argument, I remove the 13 counties with at least one oil refinery in my sample and replicate equation (4), which is shown in column (2) in table 3. Compared to the coefficients of regressions by using the whole sample, the coefficients of the sample without oil refineries decrease slightly, but the magnitude of these changes is negligible.

Second, I define fracking counties by having overlapped areas with the four main shale basins in Texas. So for counties located on the boundary of the basins, the percentages of the county area that overlapped with the basins vary greatly, shown in Figure 5. For counties with only a 5% overlapped area with a basin, although my calculation assigns them a positive number of the reserve, most of the residents in those counties might not be negatively influenced by the fracking boom. Therefore, including these counties on the boundary may downward bias the coefficients. Then I generate a new sample by dropping 42 counties that locate on the boundary of the shale plays and also have less than 50% overlapped area and replicate the regressions in table 1. The results are shown in column (3) in table 3. For asthma hospitalizations, the restricted sample almost doesn't change the coefficient that we are interested while the coefficient increase from 0.118 to 0.130 for respiratory disease hospitalizations.

Third, there exists inconsistency between energy data files from different sources. The main issue is the number of basins and shale plays they include. To generate the commuting zone level predicted oil and gas reserve per capita, I use the data of proved oil and natural gas from the Energy Information Administration (EIA), which only reports the main shale plays in the U.S.. So only the four main shale formations, Barnett Shale, Eagle Ford Shale, Haynesville Shale, and Permian Basin, have the proved oil and gas reserve data from EIA. And these four main shale formations are included in any data or document discussing Texas shales. However, the U.S. tight shale play shapefiles, which I used to calculate the overlapped area of each commuting zone with shale plays, include another small shale play called Palo Duro. Since I don't have reserve data for Palo Duro, I count those commuting zones with overlapped areas with this shale having zero reserves. Texas Railroad Commission reports another shale play called Granite Wash. The distribution of the wells on top of Granite Wash is shown in Figure 2. The three counties, Roberts county, Hemphill county, and Wheeler county, have a huge quantity of oil and gas wells scheduled for drilling. Because of the limitation of the data, I also have to assign a zero reserve to these areas in my main sample. Therefore, I remove the counties that overlapped with both Palo Duro and Granite Wash, then use the restricted sample to replicate the regressions by the standard difference-in-difference framework. These results are shown in column (4) in table 3. For asthma and respiratory disease hospitalizations, using the restricted sample doesn't change the coefficients significantly, implying that these small inconsistency issues between energy files don't influence the analysis much.

7 Explore Potential Mechanisms

7.1 Migration

There are a lot of literature empirically identifying the relationship between labor markets and migration (Black et al., 2005; Carrington et al., 1996; Kaplan and Schulhofer-Wohl, 2017; Kennan and Walker, 2011). The fracking boom generates millions of job opportunities and presents a unique opportunity to explore migration responses to positive labor demand shocks. A current paper found that fracking led to large increases in potential earnings and employment and a sizable migration response in fracking counties, and migrants to fracking areas were more likely to be male, unmarried, young, and less educated than movers more generally (Wilson, 2016). If the potential migrants moving into fracking counties have poorer respiratory health, it could also cause the observed negative impact of more respiratory disease hospitalizations. In Willis et al. (2020), North Dakota has the biggest migration response compared to other states on top of the shale basins, shown in figure 18. However, the trend of yearly numbers of inflow-migrants in Texas seems relatively flat in Wilson's figure.

To explore whether migration is a potential mechanism of the negative health impacts I found, I replicate equation (3) by replacing the Y variable of respiratory disease hospitalizations with the inflow-migrants per 100,000 residents of each county. The data of in-flow migration comes from American Community Survey (one-year estimates), and personal migration information is only available since 2005 [9] and reported in PUMA [10] level instead of county level. I transfer the data from the PUMA level to the county level first [11].

The results of the effect of the fracking boom on migration in Texas are shown in Figure 16 and Figure 17. Figure 16 shows that the coefficients of year dummies after the cutoff year are not

[9] I have to follow the staggered difference-in-difference strategy instead of the standard one since the data of in-flow migration available since 2005. Otherwise, I don't have data covered in the pre-trend period.

[10] Public Use Microdata Areas (PUMAs) are non-overlapping, statistical geographic areas that partition each state or equivalent entity into geographic areas containing no fewer than 100,000 people each. They cover the entirety of the United States, Puerto Rico, and Guam. The Census Bureau defines PUMAs for the tabulation and dissemination of decennial census and American Community Survey (ACS) Public Use Microdata Sample (PUMS) data.

[11] There exists an issue that some PUMAs consists of a few counties, and some county consists of a few PUMAs. When matching a PUMA to a county, I have to drop those PUMAs across a few different counties to keep the data accurate

significantly from zero except for one year. These results imply that compared to counties with lower reserves, counties with higher reserves in Texas don't have more inflow-migrants per capita after the fracking boom. Figure 17 presents heterogeneous analysis by age groups of migration patterns. The group of migrants between 18 to 34 years old has a decreasing pattern after four years of the fracking boom and the group of migrants older than 65 years old has an increasing pattern after two years of the fracking boom in the event study plots. The group of migrants between 0 to 17 years old and the group between 26-64 years old don't show any significant difference between counties with higher reserves and lower reserves in any year. I also estimate equation (4) by replacing the Y variable with inflow migration per capita. The results are presented in Table 4. The coefficient we are interested in for the whole sample is not statistically significant. To transfer these coefficients into a number easily understood, I still compare the county with the 25th percentile of reserve (1.282 MMBTU) and the county with the 75th percentile of reserve (46.817 MMBTU). Therefore, for residents between 18-34 years old, 45 more reserve units are equivalent to 3.92 fewer inflow migrants per 100,000 residents. For residents over 65 years old, 45 more reserve units are equivalent to 1.26 more inflow migrants. The coefficient of the latter age group is only statistically significant with a 10% confidence interval. The coefficients are not statistically significant for another two age groups (0-17 years old and 35-64 years old). These results in Table 4 are consistent to the event study patterns shown in 17. Therefore, following the staggered difference-in-difference strategy I use in this project, there is no significant difference in inflow migration between counties with higher reserves and counties with lower reserves. However, I find that fewer young adults and more people at retirement age are moving to counties with higher reserves. These results are not consistent with the findings in Wilson's paper. The potential reason might be that there are more active industries and open job positions in counties with lower reserves in Texas. However, this evidence is still enough to rule out migration as a potential mechanism that explains the negative impacts of the fracking boom on residents' respiratory health because the strongest impacts are found among non-adults. Still, there is no statistically significant migration pattern among this age group.

7.2 Water Pollution

Prior research has shown that high water contamination levels can damage health (Brainerd and Menon, 2014; Ebenstein, 2012; Lai, 2017; McKinnish et al., 2014). Shale gas development (SGD) has a life cycle involving multiple phases, including well pad preparation, drilling the well, hydraulic fracturing, and production. There are a few channels through which shale gas operations can impact water resources (Hill and Ma, 2022). Groundwater contamination could be caused in all stages of the shale gas development life cycle. The primary pathways that SGD can impact groundwater are through spills during chemical mixing and during on-site treatment and waste management, well casing failures (during fracking and through well aging), induced fractures, tank leaks, and pipeline leaks (Sun et al., 2019; Shanafield et al., 2019; Torres et al., 2016; Shrestha et al., 2017). Specifically, a study finds that drilling shale gas wells negatively impacts both drinking water quality and infant health (Hill and Ma, 2022).

Based on the current literature, water pollution caused by fracking might be a potential mechanism of the negative impacts on respiratory health found in this project. To explore this potential mechanism, I collected data files from Texas Drinking Water Watch System from 2005 to 2019 [12]. Texas Drinking Water Watch System provides detailed information on purchased groundwater samples collected in Texas, including water system name and code, county, chemical substance name and code, test method, the concentration level of specific chemical substances, and sample time. The chemical substances I include in my analysis are listed in Table 5. The raw data include more chemical substances, but there exists an issue in the Texas Drinking Water Watch System: the quality and quantity of samples that different water systems provide vary greatly. In this data, some chemical substances reported by one water system could be unavailable in other water systems. I dropped the chemical substances not reported consistently in the data and kept the 30 chemical substances with relatively higher reporting quality.

To detect the water pollution impacts of the fracking boom, I divided the 30 chemical substances into two categories, chemical substances related to fracking and chemical substances not related to fracking, according to environmental science literature (U.S. EPA, 2016; Kim et al., 2016;

[12]Texas Drinking Water Watch is a system where information about each drinking water system and detailed chemical sample details of groundwater of each water system could be accessed by querying.

Butkovskyi et al., 2017). A limitation of this data source, Texas Drinking Water Watch, is that it doesn't include all the fracking-related chemical substances. For example, 1,2,4-trimethylbenzene, benzene, bromoform, fluorene, naphthalene, and other organic chemical substances are also correlated to fracking drilling (Kim et al., 2016; Butkovskyi et al., 2017), but they are not included in the data I use. Since different chemical substances are reported by different units [13] and the magnitudes of different chemical substances are not comparable. To solve this problem, I generate an index of chemical substance contamination of the two categories of chemical substances separately by the following steps. First, I get the mean of each chemical substance in the sample. Then I divide the concentration level by the mean of the corresponding chemical substance to get a variable of the normalized concentration level of each chemical substance. Finally, I take the average of the normalized concentration levels among all the samples by each county and each year. Therefore, I get a county-year level index of water contamination for both chemical substances related to fracking and chemical substances not related to fracking.

I replicate identification equation (4)[14] by replacing the Y variable of respiratory disease hospitalizations with the water contamination indexes I generated in the last paragraph. The results are shown in Table 6. The first and second column of Table 6 presents the regression results of chemical substances related to fracking. I find that the fracking boom in Texas has increased the water contamination level. 45 more units of predicted oil and gas reserve are equivalent to 0.36 more units of water contamination index. However, the coefficients are only statistically significant with a 90% confidence interval. The third and fourth column of Table 6 presents the regression results of chemical substances unrelated to fracking, and the coefficients are not statistically significant. These findings, although not precise by the limitation of the data, are consistent with previous literature (Hill and Ma, 2022; Kim et al., 2016; Butkovskyi et al., 2017). Therefore, water pollution could be a potential source causing the negative impacts of the fracking boom on respiratory health hospitalizations in Texas.

[13]The chemical substances included in the Texas Drinking Water Watch System are reported in either UG/L or MG/L.

[14]I applied a staggered difference-in-difference strategy instead of the standard one on water pollution data since most chemical samples have been reported since either 2004 or 2005. If I use 2005 as the common cutoff year of the fracking boom, there will not be any data during the period before the fracking boom.

7.3 Air Pollution

Evidence has already been found that fracking decreases the air quality in the vicinity of an unconventional well (Zhang et al., 2020).I use NASA Terra Satellite aerosol optical depth (AOD) data[15] to estimate the effects of the fracking boom on air pollution. I pulled AOD data within circle areas with 20km as the radius of each well covered in the sample I use [16]. Then the following equation is estimated:

$$Y_{ijt} = \beta_0 + \beta_1 Treat_{ij} \times Post_{jt} + \theta X_{jt} + \gamma_j + \delta_t + \epsilon_{ijt} \tag{5}$$

In the above equation, Y_{ijt} is the AOD level in location i, neighborhood of well j, and date t. $Treat_{ij}$ is a binary measure, which equals one if location i is within a certain distance of a fracking well j. $Post_{jt}$ is another binary measure, which equals one if date t is on or after the start date of drilling of well j. X_{jt} is weather control variables of well j and date t, which includes precipitation, the maximum of the temperature, and the minimum of the temperature. γ_j and δ_t are well-fixed effects and time-fixed effects. Since I pull AOD data within circle areas with 20km as the radius and each well as the center of the circle, I use locations in the ring of 10-20km away from each well as the control group. I try different treatment groups, locations within 0-2km ring, 0-5km ring, and 0-10km ring. I also add the wind direction as the weight of the regression. The wind direction weight is measured by the cosine function value of the angle between two vectors. One is the vector from the well to the AOD reporting location, and another vector is the wind direction from the well. The bigger the angle between these two vectors, the smaller the cos function value is, which means a smaller weight in the regression.

I replicate identification equation (4)[17] by replacing the Y variable of respiratory disease hospitalizations with the water contamination indexes I generated in the last paragraph. The results

[15]Aerosal optical depth (AOD) provided by NASA's satellite is a high frequency and high resolution (3km $\times$ 3km) measure of PM.

[16]I randomly pulled a 500 unconventional well drilled between 2000 and 2014 in Eagle Ford Basin. The reason I don't use the whole sample of unconventional wells in Texas is that the size of the data covering Terra daily AOD of the entire Texas area during the study period of this project is enormous.

[17]I applied a staggered difference-in-difference strategy instead of the standard one on water pollution data since most chemical samples have been reported since either 2004 or 2005. Suppose I use 2005 as the common cutoff year of the fracking boom. In that case, there will not be any data during the period before the fracking boom.

are shown in Table 6. The results of the regressions are presented in Table 7. Panel A shows the regression results without using wind directions as weights. For different types of treatment groups, the coefficients are not statistically significant. Panel B shows the regression results with wind directions as weights. The coefficient in the first column is 1.058, which means that compared to locations within the 10-20km ring away from each well, locations in the 0-2km right has significantly higher AOD level by 1.058 units. This coefficient is statistically significant with a 95% confidence interval. When the radius of the ring of the treatment group increases, the magnitude of the coefficients decreases. As shown in column (2) and column (3), compared to locations within the 10-20km ring away from each well, locations in the 0-5km ring and 0-10km ring have significantly higher AOD levels by 0.800 unit and 0.736 unit correspondingly. These two coefficients are statistically significant with a 90% confidence interval. This evidence implies that drilling a well by fracking polluted the air in the neighborhood, and the closer the location is to the well, the poorer the air quality. Since the mean of AOD in the sample is 145.513, fracking drilling increases the AOD level within 2km of a well by 0.727%. This number is smaller than the 2.19% increased AOD found in Zhang et al. (2020). Therefore, the question of whether fracking pollutes the air in the neighborhood of the well is yes in my project, which is consistent with current literature.

8 Conclusion

This project uses inpatient hospitalization data in Texas to detect the effects of the fracking boom on the respiratory health of residents in fracking counties and finds a robust negative impact. Compared to a county with the 25th percentile of the reserves and comparing post to pre fracking periods, a county with the 75th percentile of the reserves has 0.76 (5.31) more asthma (general respiratory disease) hospitalizations per 100,000 residents, which is a 0.82% (0.55%) increase compared to the average. These impacts are more substantial when restricting the sample population to those between 0 and 17 years old than in other age groups. By back-of-envelop calculation, the fracking boom brings approximately 15,000 extra asthma hospitalizations and 103,000 respiratory disease hospitalizations to fracking counties in Texas during the period between 2005 and 2014.

I also explore the mechanisms of these findings through three different channels, migration, air pollution, and water pollution. Since I don't find significant more non-adult inflow migrants to counties with higher reserves, migration should not be the leading cause of these impacts. Instead, air and water pollution might be the potential mechanism to explain these findings. Using the NASA Terra satellite aerosol optical depth (AOD) data, fracking drilling increases the AOD level within 2km of an unconventional well by 0.727% compared to the average. I also find evidence that fracking increases the groundwater's concentration of drilling-related chemical substances. However, it is difficult to calculate the proportions of the contribution to these health impacts between the air and water pollution channels.

A limitation of this project is that inpatient hospitalization data only provide me with the opportunity to detect the lower bound of the negative impacts of the fracking boom on residents' respiratory health. The respiratory disease cases covered in the inpatient hospitalization data are those severe enough to make the inpatients get a hospital bed. Therefore, the data does not include those milder cases in my analysis. Those patients with milder symptoms may go to clinics or get over-the-counter medicines to treat themselves or do nothing if their symptoms are not severe. Therefore, the impacts of the fracking boom on respiratory health could be underestimated.

References

Black, Dan, Terra McKinnish, and Seth Sanders, "The economic impact of the coal boom and bust," *The Economic Journal*, 2005, *115* (503), 449–476.

Brainerd, Elizabeth and Nidhiya Menon, "Seasonal effects of water quality: The hidden costs of the Green Revolution to infant and child health in India," *Journal of Development Economics*, 2014, *107*, 49–64.

Busby, Christopher and Joseph J Mangano, "There'sa world going on underground—infant mortality and fracking in Pennsylvania," *Journal of Environmental Protection*, 2017, *8* (4), 381–393.

Butkovskyi, Andrii, Harry Bruning, Stefan AE Kools, Huub HM Rijnaarts, and Annemarie P Van Wezel, "Organic pollutants in shale gas flowback and produced waters: identification, potential ecological impact, and implications for treatment strategies," *Environmental science & technology*, 2017, *51* (9), 4740–4754.

Carrington, William J, Enrica Detragiache, and Tara Vishwanath, "Migration with endogenous moving costs," *The American Economic Review*, 1996, pp. 909–930.

Cascio, Elizabeth U and Ayushi Narayan, "Who needs a fracking education," *The educational response to low-skill biased technological change*, 2015, (w21359).

Chay, Kenneth Y and Michael Greenstone, "The impact of air pollution on infant mortality: evidence from geographic variation in pollution shocks induced by a recession," *The quarterly journal of economics*, 2003, *118* (3), 1121–1167.

Croquer, Aldo, David Bone, Carolina Bastidas, Ruth Ramos, and Elia García, "Monitoring coastal pollution associated with the largest oil refinery complex of Venezuela," *PeerJ*, 2016, *4*, e2171.

Currie, Janet, Matthew Neidell, and Johannes F Schmieder, "Air pollution and infant health: Lessons from New Jersey," *Journal of health economics*, 2009, *28* (3), 688–703.

_ , Michael Greenstone, and Katherine Meckel, "Hydraulic fracturing and infant health: New evidence from Pennsylvania," *Science advances*, 2017, *3* (12), e1603021.

Czolowski, Eliza D, Renee L Santoro, Tanja Srebotnjak, and Seth BC Shonkoff, "Toward consistent methodology to quantify populations in proximity to oil and gas development: a national spatial analysis and review," *Environmental health perspectives*, 2017, *125* (8), 086004.

Denham, Alina, Mary Willis, Alexis Zavez, and Elaine Hill, "Unconventional natural gas development and hospitalizations: evidence from Pennsylvania, United States, 2003–2014," *Public health*, 2019, *168*, 17–25.

Ebenstein, Avraham, "The consequences of industrialization: evidence from water pollution and digestive cancers in China," *Review of Economics and Statistics*, 2012, *94* (1), 186–201.

EPA, E.P.A U.S., "Hydraulic Fracturing for Oil and Gas: Impacts from the Hydraulic Fracturing Water Cycle on Drinking Water Resources in the United States," *Technical Report*, 2016, p. https://www.epa.gov/hfstudy.

Hill, Elaine L and Lala Ma, "Drinking water, fracking, and infant health," *Journal of Health Economics*, 2022, *82*, 102595.

Jaramillo, Paulina and Nicholas Z Muller, "Air pollution emissions and damages from energy production in the US: 2002–2011," *Energy Policy*, 2016, *90*, 202–211.

Jemielita, Thomas, George L Gerton, Matthew Neidell, Steven Chillrud, Beizhan Yan, Martin Stute, Marilyn Howarth, Pouné Saberi, Nicholas Fausti, Trevor M Penning et al., "Unconventional gas and oil drilling is associated with increased hospital utilization rates," *PloS one*, 2015, *10* (7), e0131093.

Kaplan, Greg and Sam Schulhofer-Wohl, "Understanding the long-run decline in interstate migration," *International Economic Review*, 2017, *58* (1), 57–94.

Kennan, John and James R Walker, "The effect of expected income on individual migration decisions," *Econometrica*, 2011, *79* (1), 211–251.

Kim, Seongyun, Pinar Omur-Ozbek, Ashwin Dhanasekar, Adam Prior, and Ken Carlson, "Temporal analysis of flowback and produced water composition from shale oil and gas operations: Impact of frac fluid characteristics," *Journal of Petroleum Science and Engineering*, 2016, *147*, 202–210.

Kovalenko, Alina, "Natural resource booms, human capital, and earnings: Evidence from linked education and employment records," *American Economic Journal: Applied Economics*, 2023, *15* (2), 184–217.

Lai, Wangyang, "Pesticide use and health outcomes: Evidence from agricultural water pollution in China," *Journal of environmental economics and management*, 2017, *86*, 93–120.

Logsdon, Jeanne M, *ORGANIZATIONAL RESPONSES TO*, Vol. 7, Jai Press, 1985.

Mattison, Donald R, Samuel Wilson, Christine Coussens, Dalia Gilbert et al., "Preterm Birth and Its Consequences," in "The Role of Environmental Hazards in Premature Birth: Workshop Summary" National Academies Press (US) 2003.

McKinnish, Terra, Daniel I Rees, and Peter H Langlois, "Seasonality in birth defects, agricultural production and urban location," *Economics & Human Biology*, 2014, *15*, 120–128.

Michaels, Guy, "The long term consequences of resource-based specialisation," *The Economic Journal*, 2011, *121* (551), 31–57.

Peng, Lizhong, Chad Meyerhoefer, and Shin-Yi Chou, "The health implications of unconventional natural gas development in Pennsylvania," *Health economics*, 2018, *27* (6), 956–983.

Rasmussen, Sara G, Elizabeth L Ogburn, Meredith McCormack, Joan A Casey, Karen Bandeen-Roche, Dione G Mercer, and Brian S Schwartz, "Association between unconventional natural gas development in the Marcellus Shale and asthma exacerbations," *JAMA internal medicine*, 2016, *176* (9), 1334–1343.

Rusconi, Franca, Dolores Catelan, Gabriele Accetta, Marco Peluso, Riccardo Pistelli, Fabio Barbone, Eliana Di Felice, Armelle Munnia, Paolo Murgia, Luciana Paladini et al., "Asthma symptoms, lung function, and markers of oxidative stress and inflammation in children exposed to oil refinery pollution," *Journal of Asthma*, 2011, *48* (1), 84–90.

Shanafield, Margaret, Peter G Cook, and Craig T Simmons, "Towards quantifying the likelihood of water resource impacts from unconventional gas development," *Groundwater*, 2019, *57* (4), 547–561.

Shrestha, Namita, Govinda Chilkoor, Joseph Wilder, Venkataramana Gadhamshetty, and James J Stone, "Potential water resource impacts of hydraulic fracturing from unconventional oil production in the Bakken shale," *Water research*, 2017, *108*, 1–24.

Sun, Yuqing, Di Wang, Daniel CW Tsang, Linling Wang, Yong Sik Ok, and Yujie Feng, "A critical review of risks, characteristics, and treatment strategies for potentially toxic elements in wastewater from shale gas extraction," *Environment international*, 2019, *125*, 452–469.

Torres, Luisa, Om Prakash Yadav, and Eakalak Khan, "A review on risk assessment techniques for hydraulic fracturing water and produced water management implemented in onshore unconventional oil and gas production," *Science of the Total Environment*, 2016, *539*, 478–493.

Willis, Mary, Perry Hystad, Alina Denham, and Elaine Hill, "Natural gas development, flaring practices and paediatric asthma hospitalizations in Texas," *International journal of epidemiology*, 2020, *49* (6), 1883–1896.

Wilson, Riley, "Moving to economic opportunity: the migration response to the fracking boom," *Available at SSRN 2814147*, 2016.

Zhang, Ruohao, Huan Li, Neha Khanna, DJ Sullivan, AJ Krupnick, EL Hill et al., "Satellite detection of air pollution: Air quality impacts of shale gas development in Pennsylvania," Technical Report, working paper 2020.

Figures & Tables

Figure 1: Main Shale Basins in Texas

Notes: The figure shows the four main shale basins, Permian, Barnett, Eagle Ford, and Haynesville-Bossier, that are overlapped with Texas. The figure comes from a website page: https://fingfx.thomsonreuters.com/gfx/rngs/USA-OIL-PERMIAN/010071WH3D0/index.html

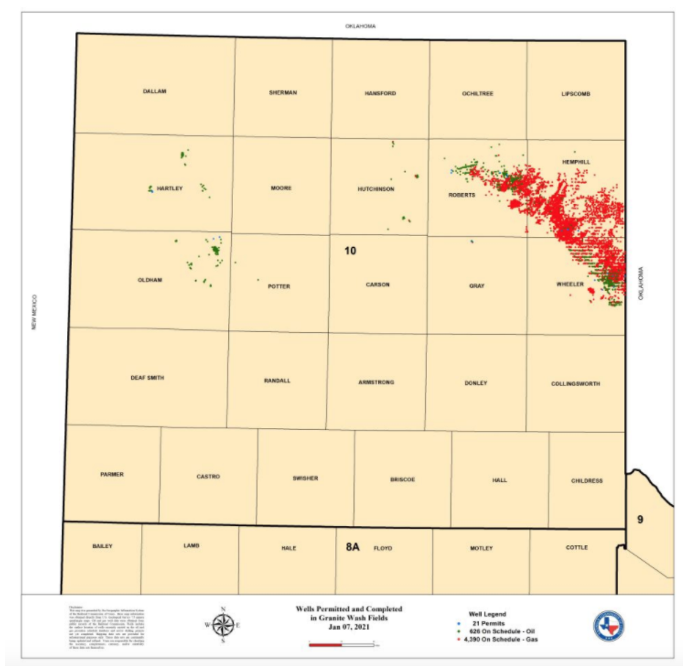
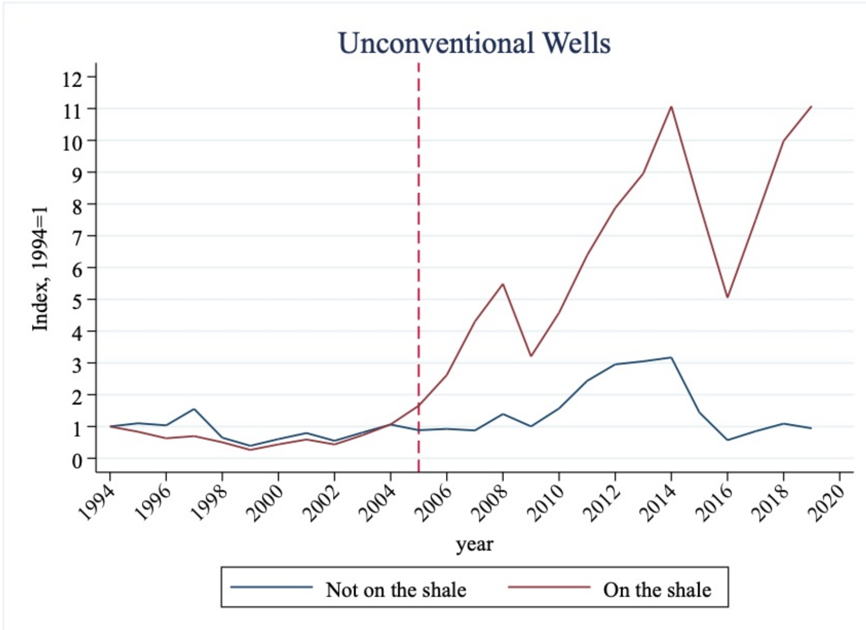


Figure 2: A shale play not included in the Data: Granite Wash

Notes: The figure shows the wells on the top of a shale play called Granite Wash, which is not included in the proved oil and gas reserve data. This figure comes from the Texas Railroad Commission website.

(a)



(b)

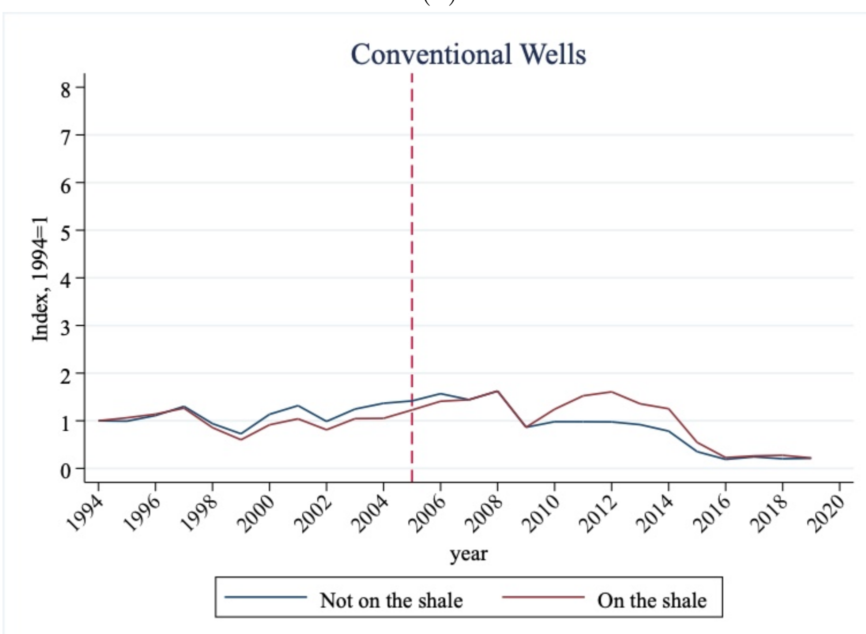
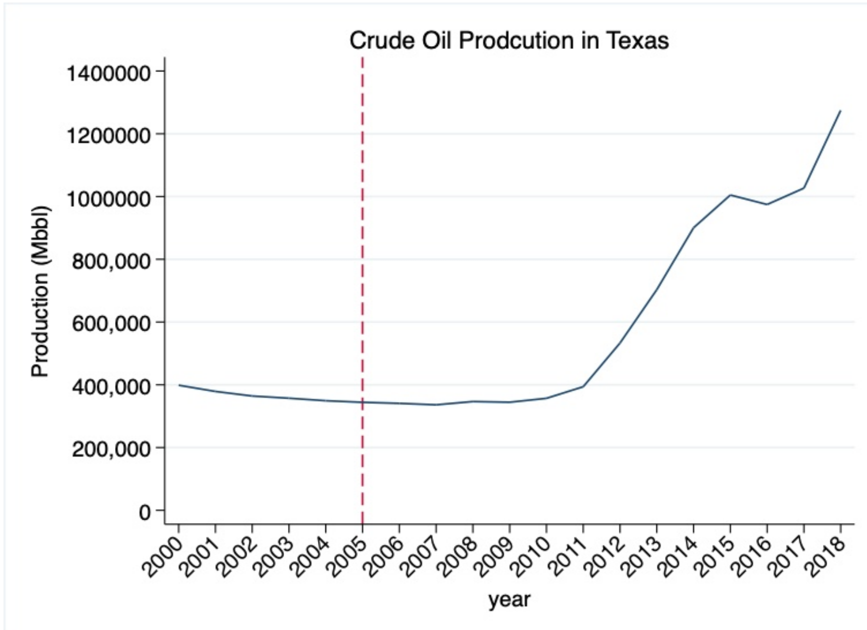


Figure 3: Oil and Gas Drilling in Texas

Notes: Panel A (B) of this figure presents the number of unconventional(conventional) wells drilled each year for counties on top of the shale and counties, not on top of the shale. This figure implies that the introduction of fracking doesn't influence traditional drilling activities too much. Still, fracking booms the drilling activities in those areas on top of the shale plays. The data used to generate these two panels is DrillingInfo.

(a)



(b)

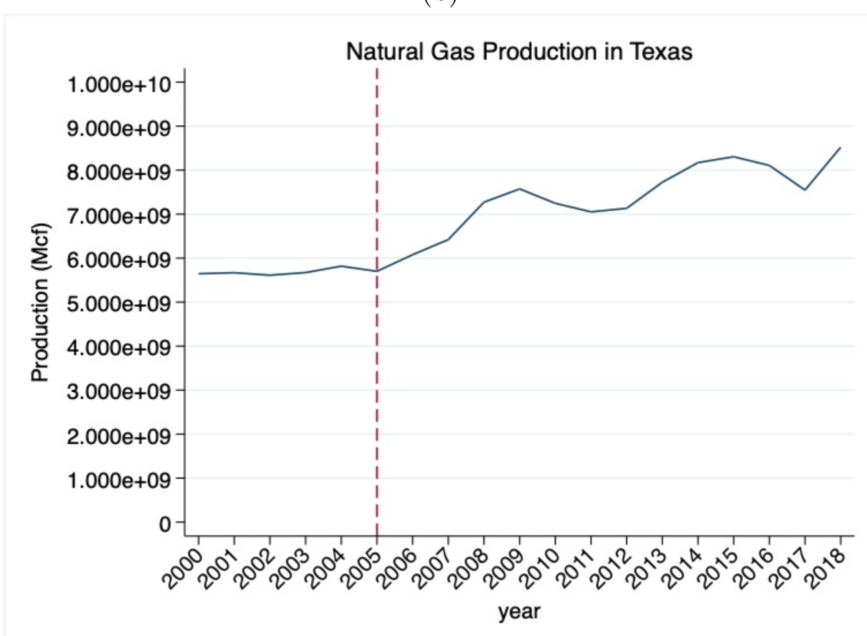


Figure 4: Oil and Gas Production in Texas

Notes: Panel A (B) of this figure presents the quantities of crude oil (natural gas) production yearly in Texas. This figure shows that Texas drastically increased crude oil and natural gas production. Especially for crude oil, the number of production tripled during the period between 2005 and 2014. The data used to generate these two figures come from the Railroad Commission of Texas.

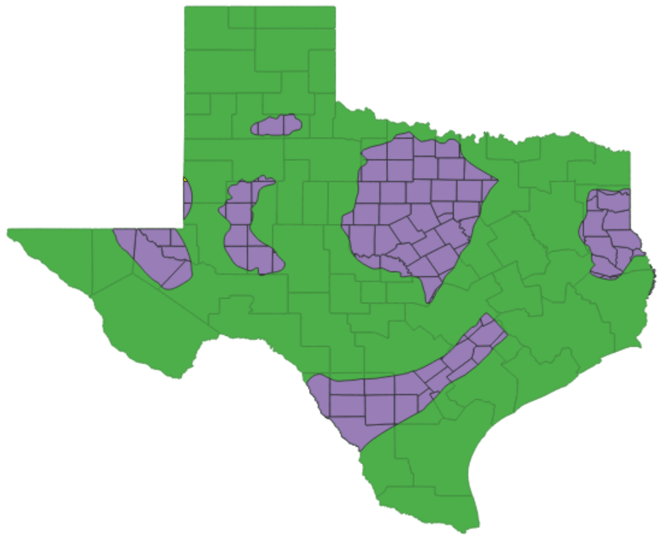


Figure 5: Overlay Basin Shapefiles and Texas Commuting Zone/County Boundary Shapefiles
Notes: The figure shows how I overlay the shapefile of the main basins in the U.S., the shapefile of the Texas commuting zone boundary, and the shapefile of the Texas county boundary. These shapefiles come from EIA and Census Bureau.

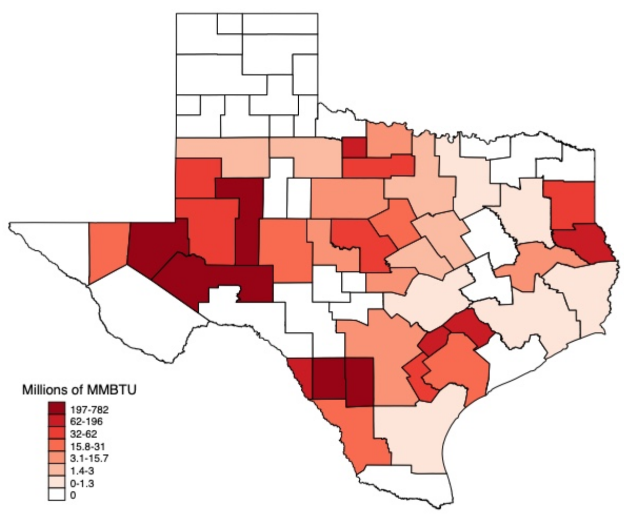


Figure 6: Distribution of Predicted Oil and Gas Reserve Per Capita in Texas
Notes: The figure plots the distribution of the predicted oil and gas reserves in Texas per capita in each commuting zone. The reserves per capita of commuting zones vary from 0 to 782 million MMBTU. The darker red a commuting zone shows in this figure, the higher reserves per capita and the higher fracking potential it has. There are clusters of highly high-reserve areas in the west and south, where the Permian Basin and Eagle Ford Basin locate. There are two other clusters of high-reserve regions found in the middle and east, where Barnett Basin and Haynesville-Bossier Basin locate.

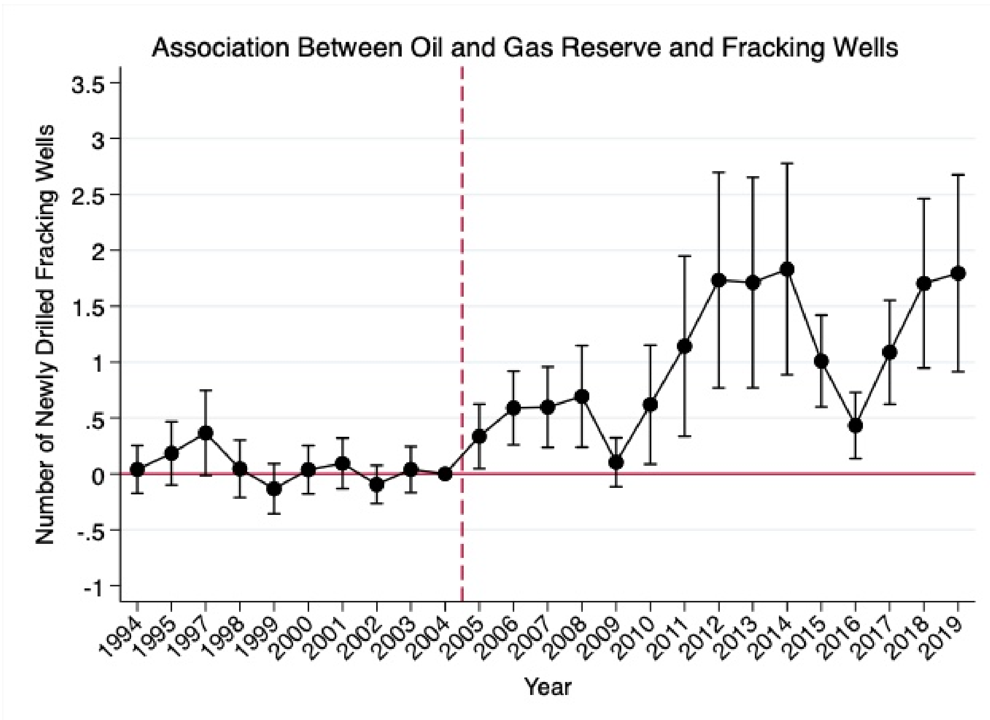


Figure 7: The Effect of the Fracking Boom on Shale Oil and Gas Drilling Activities

Notes: This figure plots the regression coefficients of newly drilled fracking wells on interactions between reserve per capita and year indicators and year and commuting zone fixed effects. As shown in this figure, commuting zones with high reserves saw significant increases in fracking wells, suggesting that underlying reserves are a good proxy for subsequent drilling activities. The number of fracking wells started to increase significantly around 2005 and peaked between 2012 and 2014. Before 2005, the coefficients are precisely estimated as zeros. This figure also shows that the starting time of the fracking boom in Texas is 2005. The data comes from DrillingInfo and EIA.

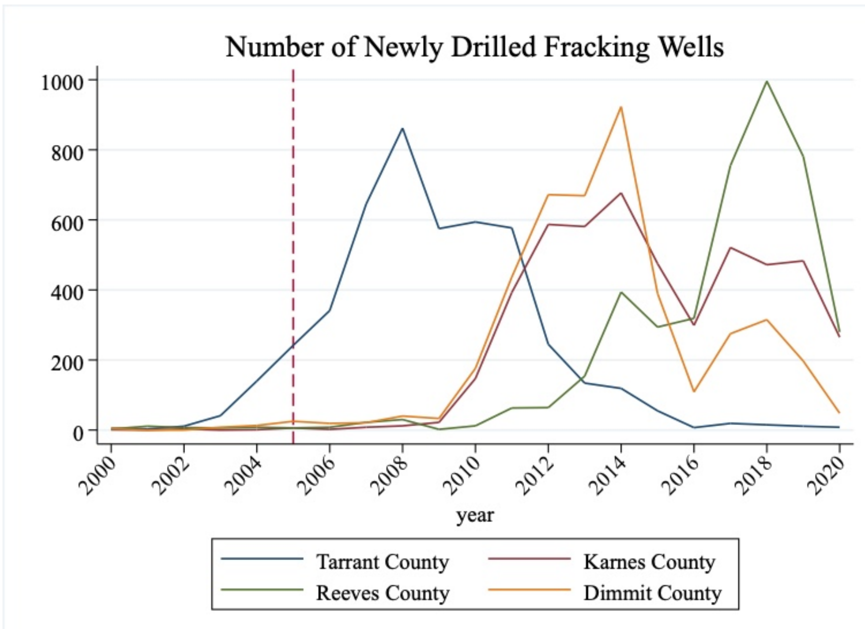
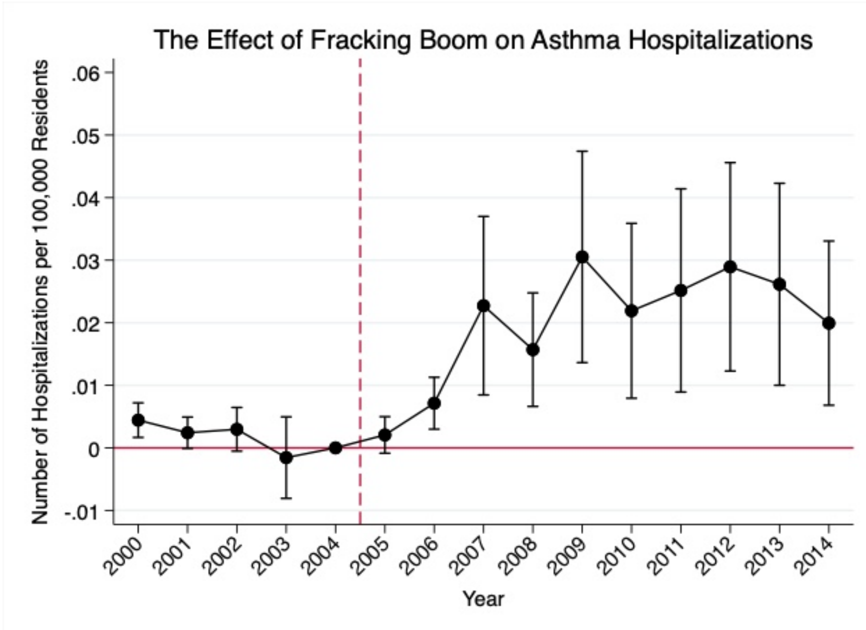


Figure 8: Numbers of Yearly Fractured Wells of 4 Largest Fracking Counties in Texas

Notes: This figure plots the number of newly fractured wells between 2000 and 2020 for the four counties with the most significant number of fracking wells in Texas. According to this figure, Tarrant county started fracking in 2002, earlier than the general starting point of the Texas fracking boom in 2005. However, Karnes county and Reeves County actively applied fracking in drilling in 2009, and Dimmit county started in 2010, which is later than in 2005. The data comes from DrillingInfo.

(a)



(b)

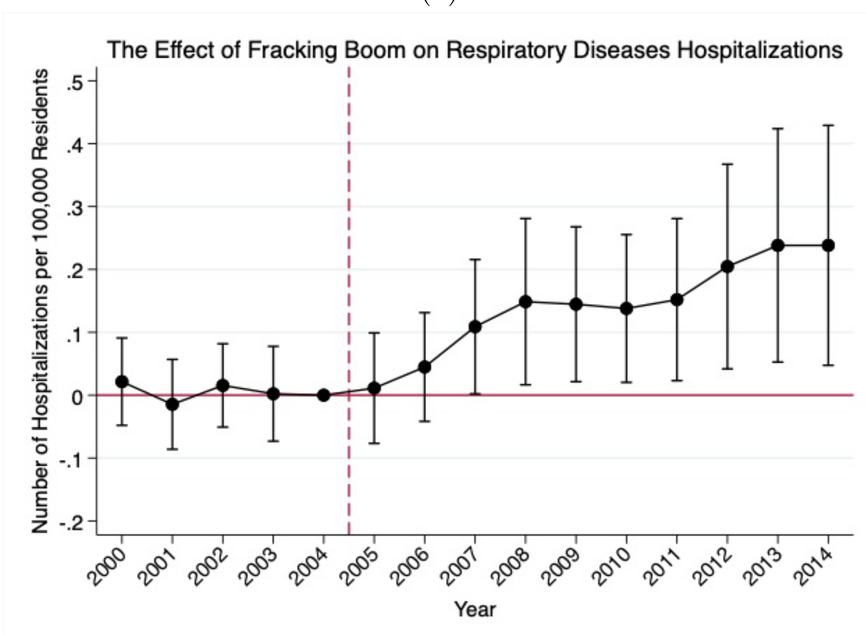
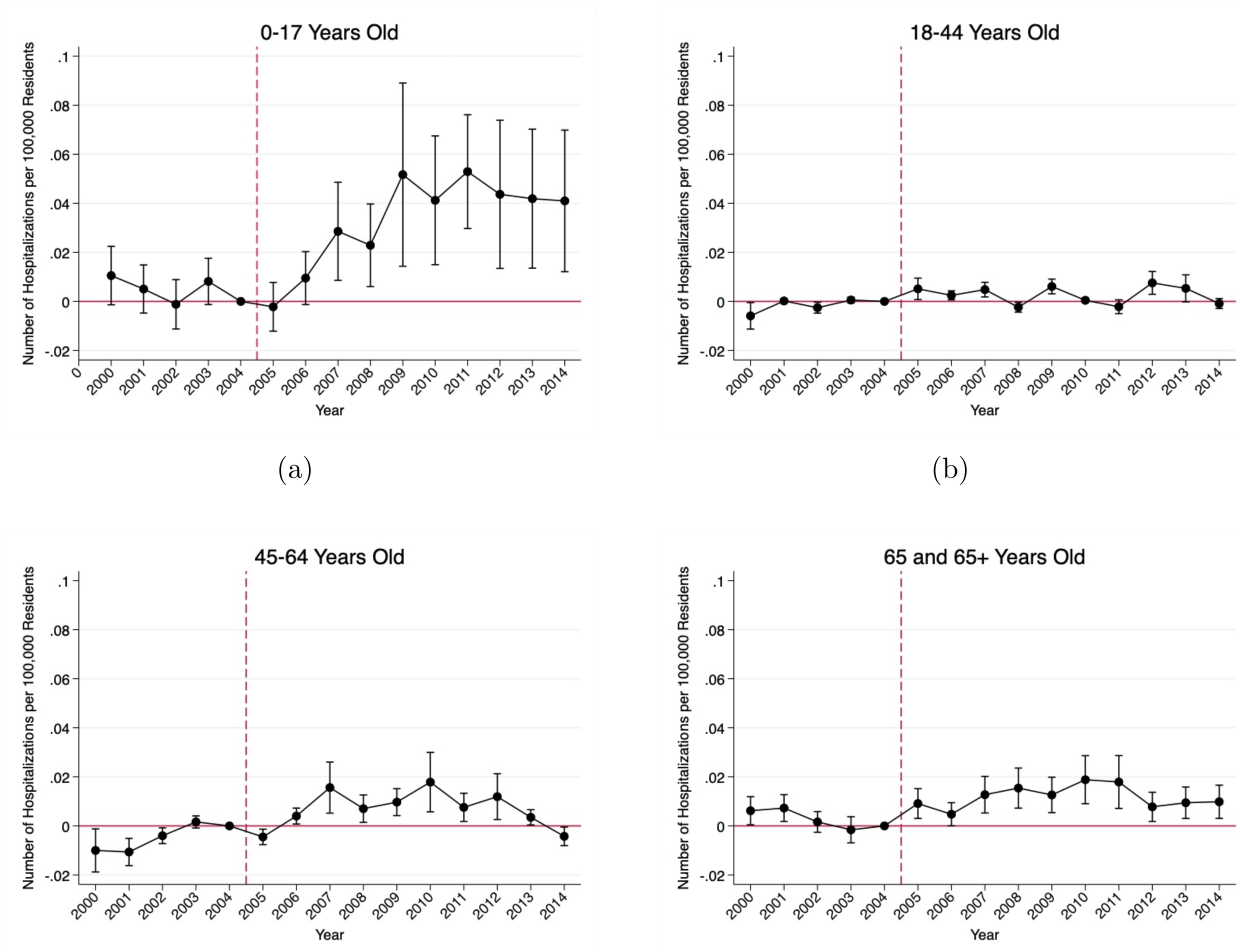


Figure 9: Effect of Fracking Boom on Hospitalizations

Notes: This figure plots the coefficients of the standard continuous difference-in-difference framework (equation (1)). Panel A shows that compared to counties with lower reserves, counties with higher reserves have significantly more asthma hospitalizations per capita since 2006 (one year after the cutoff). Panel B shows that compared to counties with lower reserves, counties with higher reserves have significantly more respiratory disease hospitalizations per capita since 2007 (two years after the cutoff). Except for the year 2000 in Panel A, the coefficients of years in the pre-trend period between 2000 and 2004 are not statistically significant from zero. These results imply that the assumption of this continuous difference-in-difference design, counties with higher reserve and counties with lower help have parallel patterns of respiratory disease hospitalizations, is generally valid.



(a) (b)

(c) (d)

Figure 10: The Effect of Fracking Boom on Asthma Hospitalizations by Age Groups

Notes: This figure plots coefficients of the standard difference-in-difference equation on asthma hospitalizations by age groups. I divide inpatients into four different age groups, 0-17 years old, 18-44 years old, 45-64 years old, and 65 and 65+ years old. All four age groups show statistically significantly more asthma hospitalizations in counties with higher reserve. The impacts are the strongest for inpatients between 0 and 17 years old, while the results are the lowest for inpatients between 18-44 years old. These four panels imply that the negative effects found of fracking on asthma are robust among different age groups, but the magnitude of the impacts varies. The fracking boom substantially impacts non-adults and people at retirement age relative to adults from 18-64 years old.

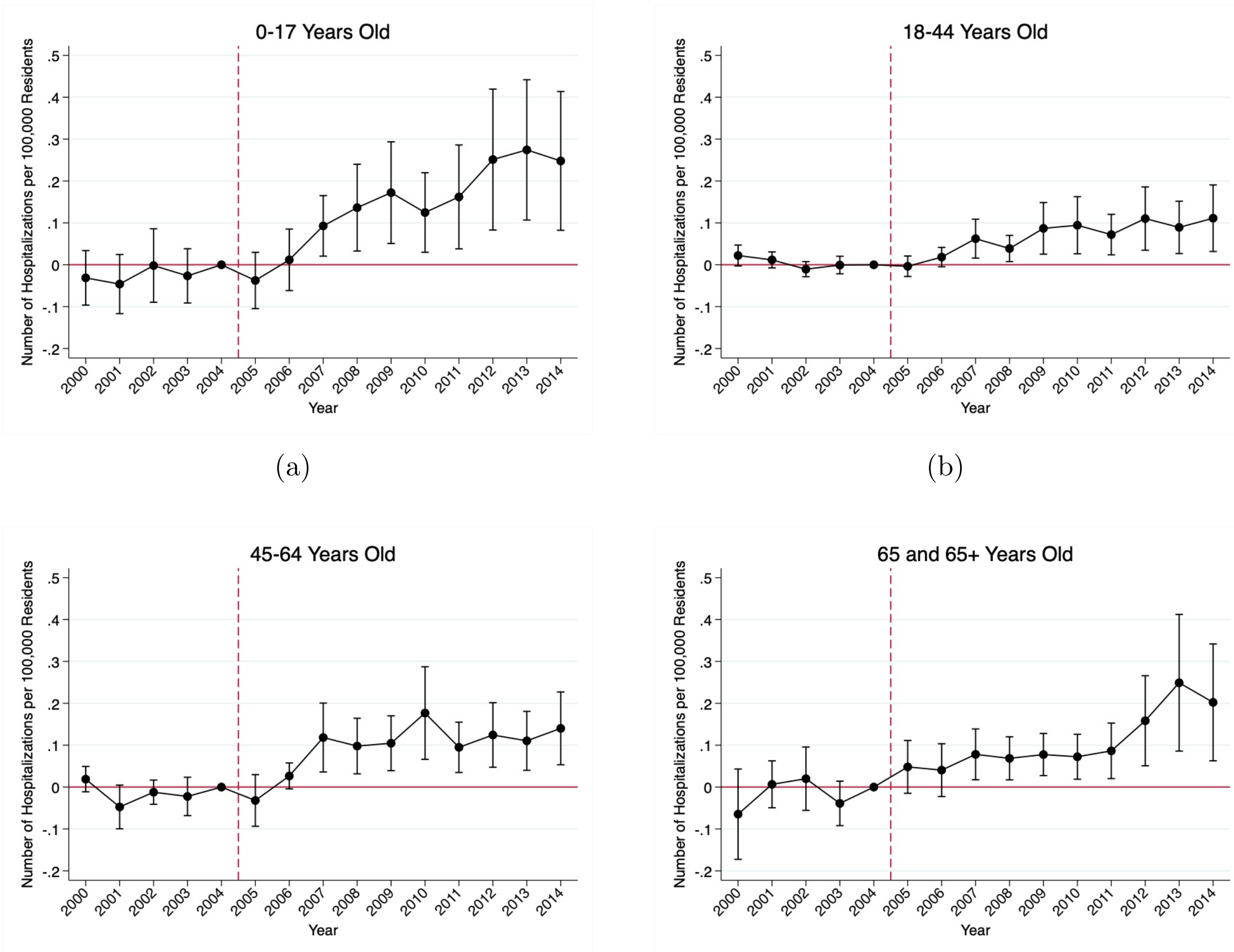
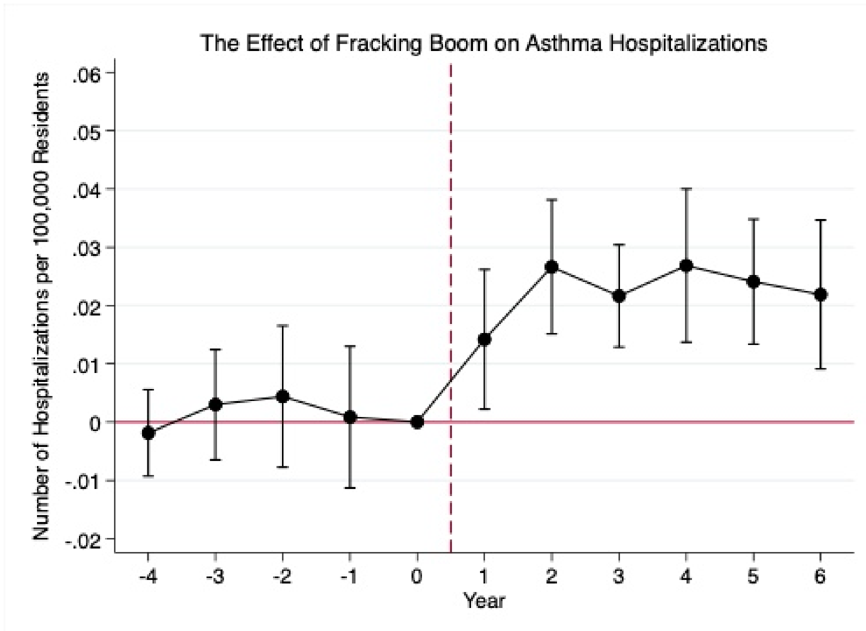


Figure 11: The Effect of Fracking Boom on Respiratory Diseases Hospitalizations by Age Groups

Notes: This figure plots the coefficients of the standard difference-in-difference equation on respiratory disease hospitalizations by age group. I divide inpatients into four different age groups, 0-17 years old, 18-44 years old, 45-64 years old, and 65 and 65+ years old. The patterns shown in this figure is similar to Figure 10.

(a)



(b)

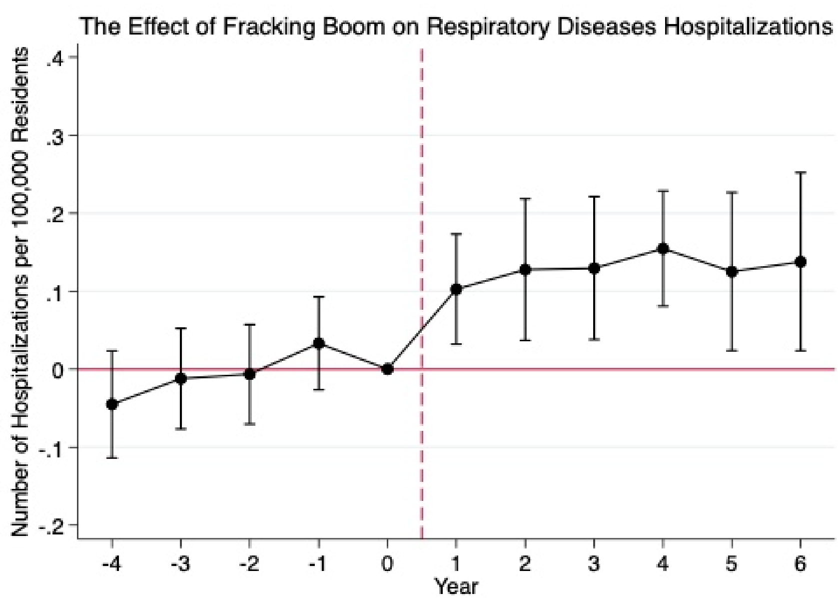
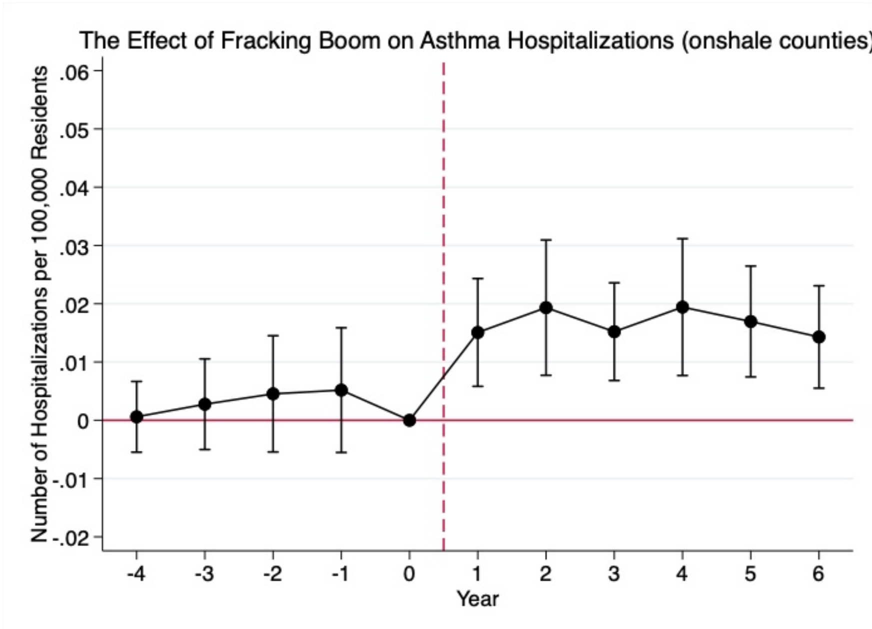


Figure 12: Effect of Fracking Boom on Hospitalizations (Staggered DiD)
Notes: This figure plots the results of the staggered difference-in-difference framework (equation (3)). According to Panel A and Panel B, asthma hospitalizations and respiratory disease hospitalizations of counties with higher reserve increase right after the cutoff year.

(a)



(b)

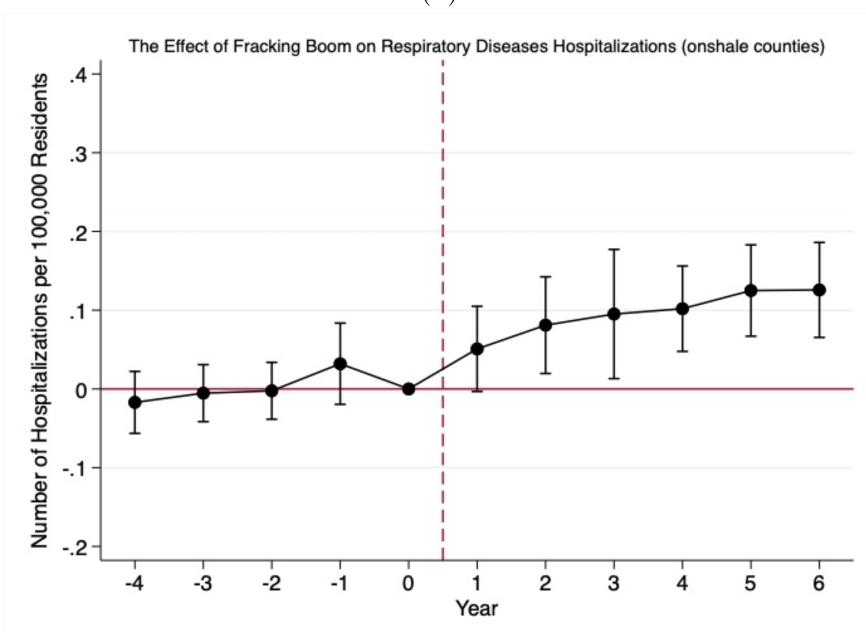


Figure 13: Effect of Fracking Boom on Hospitalizations (staggered DiD with Onshale Counties Only)

Notes:This figure plots the results of the staggered difference-in-difference framework (equation (3)) by restricting the sample to only counties on top of the shale. According to Panel A and Panel B, asthma hospitalizations and respiratory disease hospitalizations of counties with higher reserve increase right after the cutoff year.

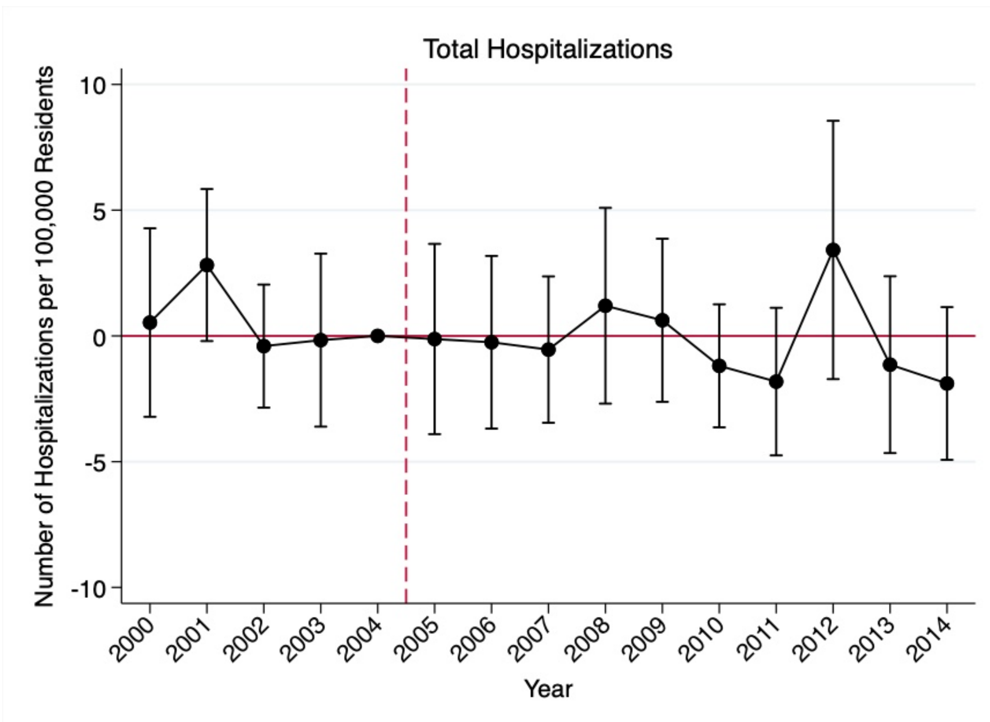
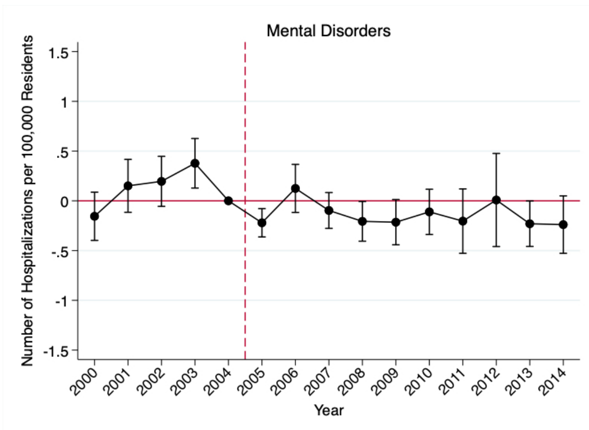
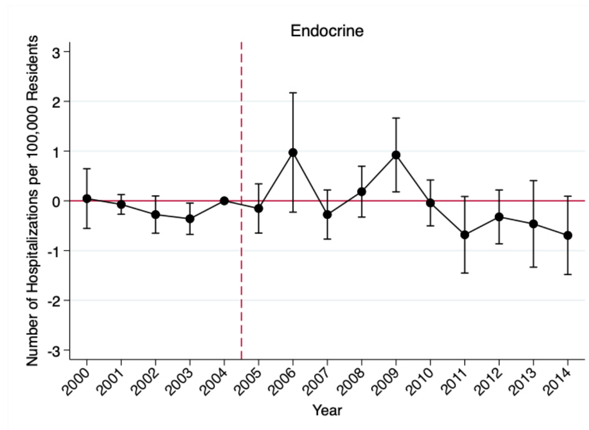


Figure 14: The Effect of the Fracking Boom on the Total Number of Hospitalizations

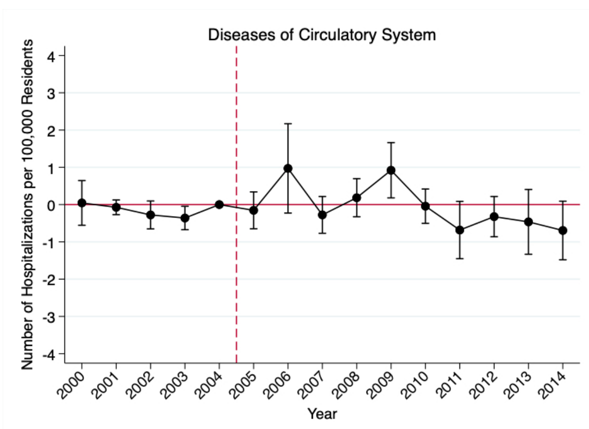
Notes: This figure plots the regressions of equation (1) by replacing the Y variable of respiratory disease hospitalizations with the total number of hospitalizations. There is no evidence that counties with higher reserves have more total hospitalizations. This figure implies that the increasing patterns of hospitalizations might be specific to respiratory tract diseases.



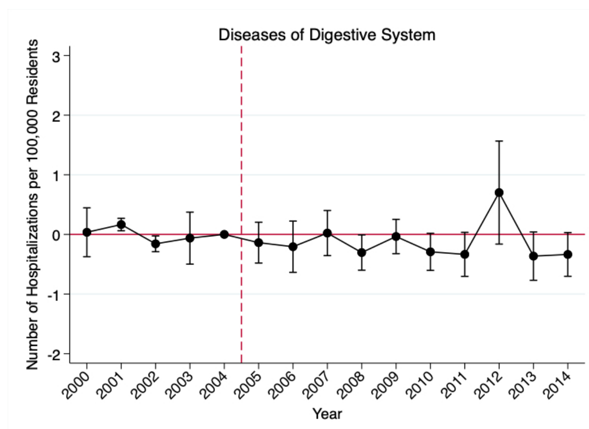
(a)



(b)



(c)



(d)

Figure 15: The Effect of Fracking Boom on Non-respiratory Disease Hospitalizations

Notes: This figure plots the event study results of four types of non-respiratory diseases, mental disorders, endocrine, diseases of the circulatory system, and diseases of the digestive system. It implies no significant patterns in non-respiratory disease hospitalizations in Texas.

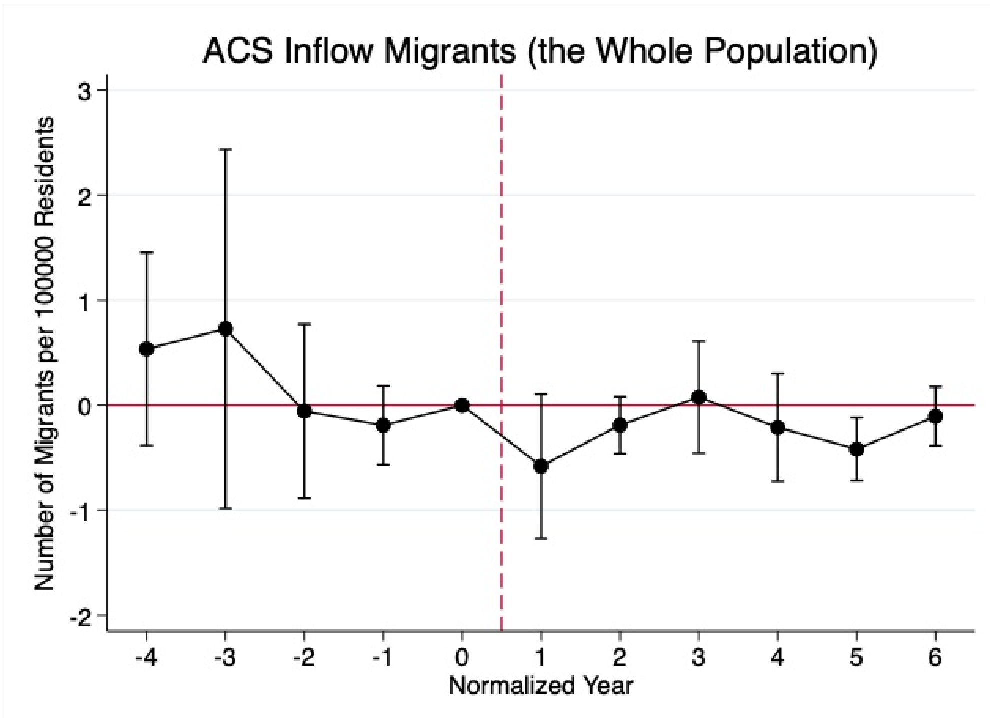
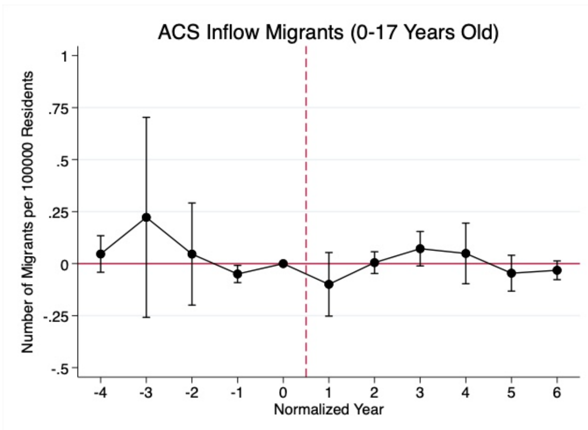
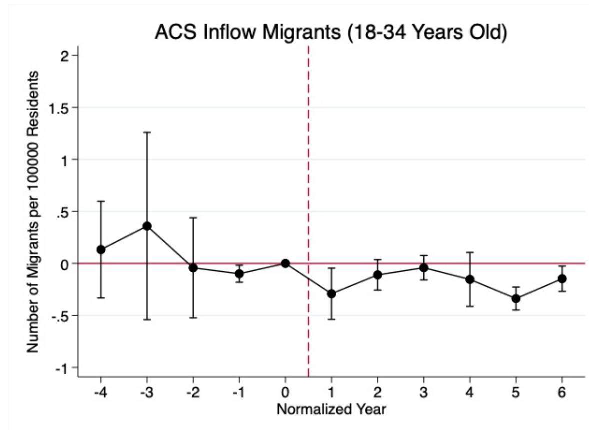


Figure 16: The Effect of the Fracking Boom on Migration

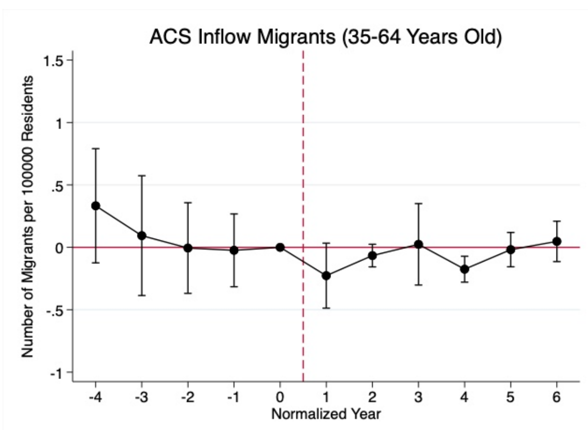
Notes: This figure plots the regression results of equation (3) by replacing the Y variable of respiratory disease hospitalizations with inflow migrants per 100,000 residents of each county. This figure shows that the coefficients of year dummies after the cutoff year are not significantly from zero except for one year, which means that compared to counties with lower reserves, counties with higher reserves in Texas don't have more inflow migrants per capita after the fracking boom. Thus, I have to follow the staggered difference-in-difference strategy instead of the standard one since the data on in-flow migration comes from American Community Survey (one-year estimates), and personal migration information is only available since 2005. Otherwise, I don't have data covered in the pre-trend period.



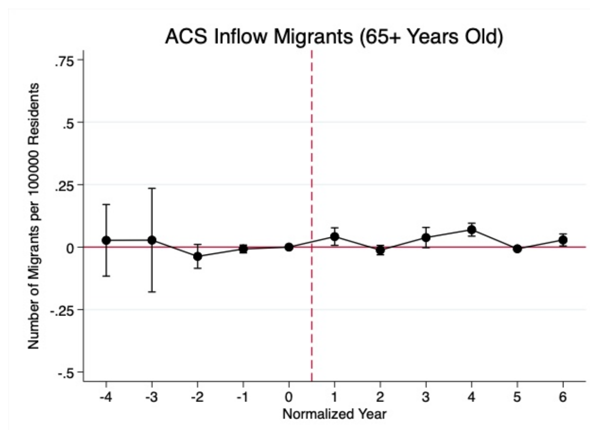
(a)



(b)



(c)



(d)

Figure 17: The Effect of Fracking Boom on Migration by Age Groups

Notes: This figure presents a heterogeneous analysis by age groups of migration patterns. Panel B shows that the group of migrants between 18 to 34 years old has a decreasing pattern after four years of the fracking boom, and Panel D shows that the group of migrants older than 65 years old has an increasing pattern after two years of the fracking boom in the event study plots. The group of migrants between 0 to 17 years old and the group between 26-64 years old don't show any significant difference between counties with a higher reserve and lower reserve in any year, shown in Panel A and Panel C in this figure.

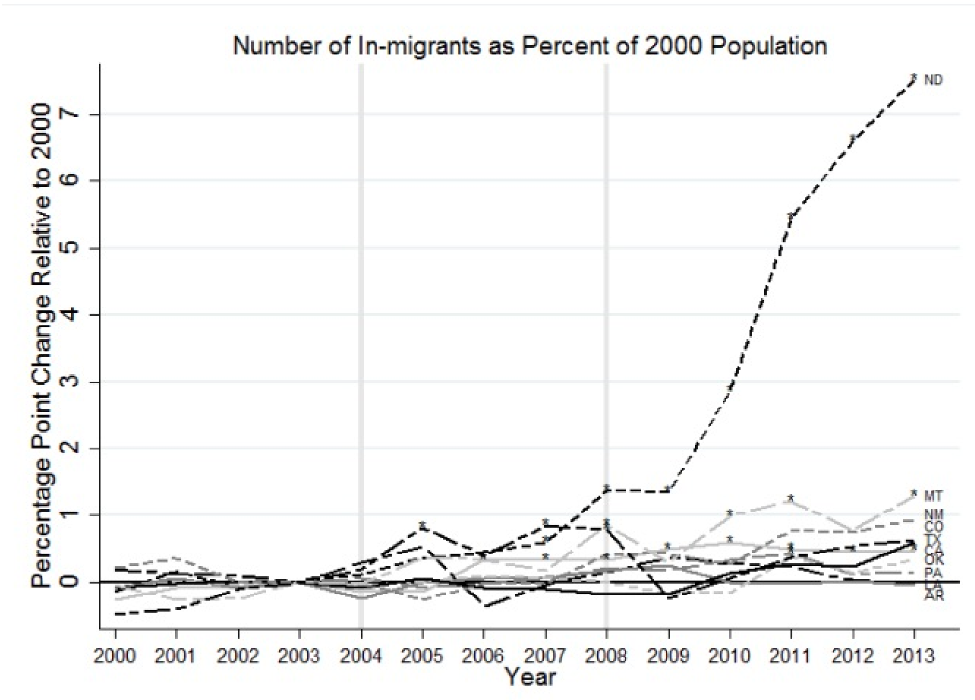


Figure 18: Trends in In-migration by State

Notes: The change in the in-migration rate for average total simulated new production in each state and year is plotted. Point estimates are obtained by regressing the in-migration rate on a set of interactions between total simulated new production between 2000 and 2013 with year indicators with county and state-by-year fixed effects. The indicator for the year 2003 is omitted as the reference year. Total simulated production is divided by the within-state average among fracking counties so that the estimated effects represent the average effect for fracking counties in that state. The vertical, gray line in 2004 and 2008 indicate the early transition years of the fracking boom. The asterisk indicates a statistically significant value at the 5 percent level.
Source: This figure comes from (Wilson, 2016) and the data comes from DrillingInfo, QWI, and IRS SOI

Table 1: The Effect of Fracking Boom on Local Residents' Hospitalizations

	(1)	(2)	(3)	(4)	(5)	(6)
	The Whole Sample		0-17	18-34	35-64	65+
Panel A: Asthma Hospitalizations						
β_1	0.016***	0.017***	0.024***	0.005***	0.009***	0.011***
	(0.006)	(0.006)	(0.008)	(0.002)	(0.003)	(0.004)
Poverty Rate		-0.042*	-0.028	-0.024*	-0.010	-0.002
		(0.730)	(0.029)	(0.014)	(0.011)	(0.012)
Log Median Household Income		-1.196	-0.343	-0.499	-0.309	-0.181
		(0.761)	(0.742)	(0.458)	(0.309)	(0.271)
CZ Fixed Effect	×	×	×	×	×	×
Year Fixed Effect	×	×	×	×	×	×
N of Observations	3507	3507	3154	2525	2931	2770
Panel B: Respiratory Hospitalizations						
β_1	0.117***	0.118***	0.146***	0.045***	0.078***	0.097***
	(0.039)	(0.039)	(0.048)	(0.014)	(0.028)	(0.029)
Poverty Rate		1.615	0.808	0.131	0.445	0.853
		(1.097)	(0.583)	(0.061)	(0.415)	(0.540)
Log Median Household Income		43.103	19.473	2.439	5.150	13.961
		(40.279)	(20.465)	(1.714)	(9.320)	(14.259)
CZ Fixed Effect	×	×	×	×	×	×
Year Fixed Effect	×	×	×	×	×	×
N of Observations	3784	3784	3682	3432	3716	3768

Notes: This table presents the regressions of equation (1) by restricting the sample to different age groups. Column (1) and (2) shows the results of baseline regressions by the whole sample. Column (2) to (6) shows the regression results of the samples by population age among 0-17 years old, 18-34 years old, 35-64 years old, and 65+ years old. This table implies that the negative impacts found of fracking on asthma are robust among different age groups, but the magnitude of the effects varies. For example, the fracking boom substantially impacts non-adults and people at retirement more than adults from 18-64 years old. Significance: *** at 1 percent level, ** at 5 percent level, * at 10 percent level.

Source: Texas Inpatient Public Use Data Files, DrillingInfo and EIA

Table 2: The Effect of Fracking Boom on Local Residents' Hospitalizations (Robustness Checks)

	(1)	(2)	(3)	(4)
	Whole Sample	Without Refineries	Without Boundary	Without Small Shales
Panel A: Asthma Hospitalizations				
β_1	0.017***	0.016***	0.018***	0.016***
	(0.006)	(0.006)	(0.006)	(0.006)
Poverty Rate	0.042*	-0.021	-0.02	-0.04
	(0.730)	(0.029)	(0.026)	(0.029)
Log Median Household Income	-1.196	-0.083	-0.225	-0.659
	(0.761)	(0.747)	(0.617)	(1.024)
CZ Fixed Effect	×	×	×	×
Year Fixed Effect	×	×	×	×
N of Observations	3507	3312	2925	3363
Panel A: Respiratory Hospitalizations				
Beta_1	0.118***	0.114***	0.130***	0.114***
	(0.039)	(0.038)	(0.042)	(0.038)
Poverty Rate	1.615	0.967	1.221	2.101
	(1.097)	(1.239)	(0.827)	(1.353)
Log Median Household Income	43.103	31.721	28.257	76.71**
	(40.279)	(39.33)	(39.33)	(37.781)
CZ Fixed Effect	×	×	×	×
Year Fixed Effect	×	×	×	×
N of Observations	3784	3589	3184	3604

Notes: This table presents the regressions of robustness checks described in section 6.2. Column (1) shows the results of baseline regressions by the whole sample. Column (2) shows the results of a restricted model by removing the 13 counties with at least one oil refinery. Column (3) shows the results of a restricted sample by dropping 42 counties that locate on the boundary of the shale plays and also have less than 50% overlapped area. Finally, column (4) shows the results of another restricted sample by removing the counties that overlapped with Palo Duro and Granite Wash and two smaller shale plays not covered in my reserve data. Compared to the coefficients in column (1), the coefficients of the restricted samples vary a little bit, but the magnitude of these changes is negligible. Significance: *** at 1 percent level, ** at 5 percent level, * at 10 percent level.

Source: Texas Inpatient Public Use Data Files, DrillingInfo and EIA

Table 3: The Effect of Fracking Boom on Non-respiratory Disease Hospitalizations

	(1)	(2)	(3)	(4)	(5)	(6)
	Total Hospitalizations		Mental Disorders	Endocrine	Circulatory	Digestive
β_1	-0.384	-0.370	-0.239	-0.024	0.117	-0.143
	(1.147)	(1.149)	(0.249)	(0.156)	(0.188)	(0.169)
Poverty Rate		19.866	2.757	-0.345	-29.352**	-2.824
		(52.073)	(4.437)	(2.7620	(12.237)	(4.290)
Log Median Household Income		-496.358	121.098	-32.377	-742.118**	-124.687
		(1550.459)	(102.613)	(93.779)	(331.664)	(135.349)
CZ Fixed Effect	×	×	×	×	×	×
Year Fixed Effect	×	×	×	×	×	×
N of Observations	3,807	3,807	3,758	3,778	3,805	3,794

Notes: This figure shows replicated equation (2)' results by using total hospitalizations and restricted samples of several categories of non-respiratory tract disease hospitalizations. However, the coefficients of the total hospitalizations and the four samples of non-respiratory diseases, mental disorders, endocrine, circulatory system diseases, and digestive system diseases are not statistically significant.

Source: Texas Inpatient Public Use Data Files, DrillingInfo and EIA

Table 4: The Effect of Fracking Boom on Migration

	(1)	(2)	(3)	(4)	(5)	(6)
	The Whole Sample		0-17	18-34	35-64	65+
β_1	0.087	0.079	0.017	-0.087***	0.032	0.028*
	(0.060)	(0.056)	(0.016)	(0.028)	(0.023)	(0.0160)
Poverty Rate		1.092	0.247	-0.679	0.405	-0.274
		(3.731)	(0.689)	(1.722)	(1.322)	(0.854)
Log Median Household Income		36.293	11.195	-3.207	6.162	5.019
		(127.634)	(22.369)	(48,784)	(44.636)	(26.791)
CZ Fixed Effect	×	×	×	×	×	X
Year Fixed Effect	×	×	×	×	×	×
N of Observations	967	967	967	967	967	963

Notes: This table presents the estimates of equation (4) by replacing the Y variable with inflow migration per capita. Column (1) and column (2) show the results of the whole sample, and the coefficient we are interested in is not statistically significant. Column (3) to (6) shows the results of the four age groups correspondingly. To transfer these coefficients into a number easily understood, I still compare the county with a 25 percentile of reserve (1.282 MMBTU) and the county with a 75 percentile of reserve (46.817 MMBTU). Therefore, for residents between 18-34 years old, 45 more units of the reserve are equivalent to 3.92 fewer inflow migrants per 100,000 residents, and for residents more than 65 years old, 45 more units of the reserve are equivalent to 1.26 more inflow migrants. The coefficient of the latter age group is only statistically significant with a 10% confidence interval. The coefficients are not statistically significant for another two age groups (0-17 years old and 35-64 years old). These table results are consistent with the event study patterns shown in 17.
Source: American Community Survey

Table 5: Chemical Substances Related and Not Related to Fracking

Chemical Substance	Fracking Related	Chemical Substance	Fracking Related
ALKALINITY, TOTAL	No	MAGNESIUM	No
BROMODICHLOROMETHANE	No	MANGANESE	No
BROMOFORM	No	MONOBROMOACETIC ACID	No
CALCIUM	No	MONOCHLOROACETIC ACID	No
CARBON DISULFIDE	Yes	NITRITE	No
CARBON, TOTAL	Yes	SILVER	No
CHLORIDE	Yes	SODIUM	Yes
DIBROMOACETIC ACID	No	SULFATE	No
DIBROMOCHLOROMETHANE	No	THALLIUM, TOTAL	No
DICHLOROACETIC ACID	No	TOTAL HALOACETIC ACIDS (HAA5)	No
ETHYLBENZENE	Yes	TRICHLOROACETIC ACID	No
FLUORIDE	No	TTHM	No
HARDNESS, TOTAL (AS CACO3)	Yes	XYLENE, META AND PARA	Yes
IRON	No	XYLENES, TOTAL	Yes
LEAD	No	ZINC	No

Notes: These chemical substances come from the reports of Texas Drinking Water Watch Systems. The raw data include more chemical substances, but there exists an issue that the quality and quantity of samples provided by different water systems vary a lot. Therefore, I dropped the chemical substances not reported consistently in the data and kept the 30 chemical substances with relatively higher reporting quality. A limitation of this data source, Texas Drinking Water Watch, is that it doesn't include all the fracking-related chemical substances. 1,2,4-trimethylbenzene, benzene, bromoform, fluorene, naphthalene, and other chemical substances are also correlated to fracking drilling (Kim et al., 2016; Butkovskyi et al., 2017). However, the data I use only covers the eight chemical substances related to fracking, which are listed in the above table.

Source: Texas Drinking Water Watch

Table 6: The Effect of Fracking Boom on Water Pollution

	(1)	(2)	(3)	(4)
	Fracking Related		Fracking not Related	
β_1	0.008*	0.008*	0.009	0.009
	(0.005)	(0.005)	(0.007)	(0.007)
Poverty Rate		-0.002		0.001
		(0.008)		(0.004)
Log Median Household Income		-0.283		-0.007
		(0.204)		(0.146)
CZ Fixed Effect	×	×	×	×
Year Fixed Effect	×	×	×	×
N of Observations	376	376	886	886

Notes: This table presents the regression results of equation (4) by replacing the Y variable of respiratory disease hospitalizations with the water contamination indexes I generated in Section 7.2. The first and second column of this table presents the regression results of chemical substances related to fracking. I find that the fracking boom in Texas has increased the water contamination level. Forty-five more units of predicted oil and gas reserve are equivalent to 0.36 more units of water contamination index. However, the coefficients are only statistically significant with a 90% confidence interval. The third and fourth columns present the regression results of chemical substances unrelated to fracking, and the coefficients are not statistically significant.
Source: Texas Drinking Water Watch

Table 7: The Effect of Fracking Boom on Air Pollution

	(1)	(2)	(3)
	0-2km	0-5km	0-10km
Panel A: Unweighted Regression			
Treatment effect	1.956	1.431	1.058
	(1.718)	(1.015)	(0.787)
Weather Control Variables	×	×	×
Well and Time Fixed Effects	×	×	×
Adjusted R2	0.43	0.42	0.42
N	46,960,128	58,700,241	78,267,594
Panel B: Weighted Regression			
Treatment effect	1.058**	0.800*	0.736*
	(0.516)	(0.457)	(0.408)
Weather Control Variables	×	×	×
Well and Time Fixed Effects	×	×	×
Adjusted R2	0.25	0.37	0.39
N	46,960,128	58,700,241	78,267,594

Notes: This table presents the results of the regressions in equation (5), which are the effects of the fracking boom on air pollution. Panel A shows the regression results without using wind directions as weights. For different types of treatment groups, the coefficients are not statistically significant. Panel B shows the regression results with wind directions as weights. The coefficient in the first column is 1.058, which means that compared to locations within the 10-20km ring away from each well, areas in the 0-2km right has significantly higher AOD level by 1.058 units. This coefficient is statistically significant with a 95% confidence interval. Suppose we increase the radius of the treated ring, the magnitude of the coefficient decreases. Shown in column (2) and column (3), compared to locations within the 10-20km ring away from each well, areas in the 0-5km ring and 0-10km ring have significantly higher AOD levels by 0.800 unit and 0.736 unit correspondingly. These two coefficients are statistically significant with a 90% confidence interval. This evidence implies that drilling a well by fracking polluted the air in the neighborhood, and the closer the location is to the well, the poorer the air quality. Since the mean of AOD in the sample is 145.513, fracking drilling increases the AOD level within 2km of a well by 0.727%.

Source: NASA Terra Satellite AOD Data