THESIS

ECONOMIC EFFECTS OF OIL AND GAS DEVELOPMENT ON CHILD HEALTH IN COLORADO

Submitted by

Susan S Ekoh

Department of Agriculture and Resource Economics

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Spring 2015

Master's Committee:

Advisor: Andrew Seidl,
Co-Advisor: Jordan Suter

Anita Pena

ABSTRACT

ECONOMIC EFFECTS OF OIL AND GAS DEVELOPMENT ON CHILD HEALTH IN COLORADO

The oil and gas industry is a huge contributor to the economy of many countries and states worldwide. Even though the industry creates jobs and income, debates continue surrounding the environmental and human costs of the industry. There are many health concerns in particular because studies have shown that there is the potential for water and air pollution from emissions that are generated through oil and gas production processes.

This study aims to contribute to the body of knowledge on this issue by exploring the potential economic effects of oil and gas development on child respiratory health. Using state in-patient data from the Agency for Research and Health Quality (ARHQ)-Healthcare Cost and Utilization Project (HCUP), I estimate difference-in-difference results to measure the potential effects of oil and gas production in Colorado over time, between the years 2004 and 2013 and for children between ages 0 and 19.

Results reveal that over time from 2004 until 2013, oil and gas had no significant effect on child respiratory health. For individual years, results for 2013 show significant effects. The overall lack of effect of oil and gas development on child respiratory health from this study could be that the respiratory cases reported were as a result of factors other than air pollution from oil and gas activities.

ACKNOWLEDGEMENTS

I would like to thank all of my advisors and committee members Dr. Andy Seidl, Dr. Jordan Suter and Dr. Pena for their kind and valuable inputs during the period of this research. Also thanks to Dr. Cutler for making out time to be a part of my defense. I also appreciate the help of Elaine Hill who provided insights to available data sources.

I would like to thank Dr. Suter particularly for being kind and willing to offer financial support to help purchase the data and for helping me see the big picture regarding this research.

To the African Development Bank for providing funds for my tuition through this entire degree and finally, to my friends and family for their support and prayers.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Air and water pollution are important contributors to public health outcomes. A number of studies have been carried out to show the effects of pollution on human health (Sebastian et al.2001; Janjua, et.al 2006; Hall et al. 2010; Colborn, et.al 2011; Hill 2012). Many studies on the effects of oil and gas production emissions involve the collection of air and water samples, laboratory testing of these samples and then a final process of comparing results to expected or recommended air and water quality measures. (McKenzie et al. 2012; Currie et al. 2009; Allen et al, 2013; Ferer et al, 2013). There are several sources of these types of pollution: exhaust pollution from vehicles, industrial pollution, etc. Oil and gas production has been identified as a contributor to water and air pollution through the processes involved such as exploration, drilling, refining, and transportation (Rozell and Reaven, 2012; O'Rourke and Connolly, 2003). These result in costs to the environment, human, animal and aquatic populations/systems. And although the benefits from the oil and gas industry cannot be disputed, in order to make appropriate policy decisions it is important to address the major costs and trade-offs involved.

This paper therefore, focuses on estimating the human health effects of oil and gas production, particularly through air pollution using respiratory incidences among the study population and hence contributes to existing knowledge available on this problem. The overall objective is to improve understanding, and provide viable contributions for policy making decisions related to oil and gas exploration.

For this study, I utilize data on state inpatient respiratory cases from the Healthcare Cost Utilization Project (HCUP) in the Agency for Healthcare Research and Quality for the years 2004 and 2013. Data on producing oil and gas wells in Colorado are also used form the Colorado Oil and Gas Commission as wells as data from the US Census Bureau for demographic data. I run an econometric analysis using a difference-in-difference approach which is discussed further in the paper, estimating the potential effects of oil and gas development on respiratory illnesses in the state of Colorado over time.

CHAPTER 2: LITERATURE REVIEW

2.1 Human Health Effects of Oil and Gas Production

Water and air pollution are an important source of illness to humans depending on the level/degree of exposure. Oil and gas production has been found to be a contributor to water and air pollution due to emissions of pollutants that occur during the production process. Witter et al. (2008) report a few studies that have examined the effects of oil and gas activities on air and water quality in the Western slope of Colorado. Oil and gas production greatly affects the environment and humans in general, and human health can be negatively affected due to excessive exposure to harmful emissions and also through spills that occur through the production and transportation processes (Colborn, et.al 2011; O'Rourke and Connolly, 2003).

Several research studies have addressed the potential effects of oil and gas production on humans and the environment. These studies are typically carried out after an oil spill. Although the aim of this paper is to analyze the economic implications of such health effects in the state of Colorado, it is important to establish a causal relationship between oil and gas production and pollution and finally, its effects on human health. This section of the paper will therefore explore the epidemiological literature and the second part will look at the economic implications.

Hydraulic fracturing has been one of the major sources of debate as regards the potential harmful implications of the natural gas and oil production process. Hydraulic fracturing is one of the common methods of natural gas drilling in many regions of the United States of America. It is "a drilling technique that improves access to unconventional natural gas within the miniscule pores of shale deposits" (Soeder, 2010). It was developed in the late 1940s to reach natural gas deposits which are not easily accessible and which are found within underground coal and shale pores

(Finewood and Stroup, 2012). This technique is more popular today because of its relative cost effectiveness and its ability to access a wider range of deposits (Meyer et al, 2010).

One of the major concerns over the hydraulic fracturing process is that it is heavily dependent on the use of "undisclosed types and amounts of chemicals" as wells as drilling fluids containing heavy metals that could contaminate water resources (Doyle, 1994; Horwitt, 2009; Osbourn et al. 2011) and in the release of gases into the atmosphere that are potentially harmful to human populations.

Hydraulic fracturing (or 'fracking') has been a source of huge debate as to the appropriate policies to be put in place in order to reduce exposure to possible illness causing pollution sources. It has thus, raised a lot of health concerns on the adverse effects it has on human health through exposure to emissions from air and water. The natural gas industry argues that fracking is not so much of a risk since ground water occurs at a shallower depth than the sources of shale gas deposits (Schmidt, 2011). A recent report, however, shows that 'the Colorado Oil and Gas Conservation Commission recorded a total number of 1549 spills within a four year period (Jan 2003-March 2008)". And at the time of this study (2014) the COGCC recorded a total number of 173 spills with in the first quarter of 2014 (COGCC).

Human health can also be affected through increased lung infections, respiratory illnesses, skin infections, etc. Respiratory and lung infections are often as a result of air pollution while skin infections are most likely due to contact with polluted water. Some studies have also shown that there is increased risk of diseases from residing in close proximity to these areas or sources of air pollution due to exposure to harmful pollutants. Specifically, residents within less than half a mile have been shown to be at a higher risk than those within greater than half a mile to 5 miles of an oil well (McKenzie et. al, 2012).

According to Witter et al. (2008), The Earthworks Oil and Gas Accountability project conducted a study of the effects of Marcellus Shale development in Central New York and Pennsylvania. After administering surveys and conducting air and water tests, they find that people residing closest to the gas facilities had a higher percentage of respiratory and sinus complaints.

Hill (2012) uses a difference-in-difference approach in looking at the effects of natural gas development on infants in Pennsylvania. This difference-in-difference approach helps to identify before and after effects. That is, it helps to identify a causal relationship between a treatment and its resulting effect while comparing it to a control without the treatment. This is discussed further in section 3.1 of the paper. In this case, she looks at birth outcomes of infants before and after a well was completed. The birth outcomes of interest in the analysis were: birth weight, the five APGAR[1] scores, premature birth and small for gestational weight. Her results show that exposure to natural gas development resulted in poorer infant health outcomes**.**

2.2 Cases and Studies of Health Effects of Oil and Gas around the World

2.2.1 Epidemiological Approaches to Health Effect Studies

One of the ways in which the potential health effects of oil and gas development has been tested is through the application of the oil to the skin of mice. These studies found a higher than expected incidence of skin tumors after application (Grimer et al, 1982, Wilson et al.1988, Clark et al.1988) suggesting that oil and gas have harmful chemicals that could lead to adverse effects

[1] An APGAR Score is a score system used to represent the conditions of an infant after birth, it measures Appearance, Pulse, Grimace, Activity and Respiration

on health. However we cannot conclude that the same effect on mice will apply to humans without concrete evidence (International Agency for Research on Cancer, 1989). Another study has tried to show the potential effects of cancer from exposure to contaminating chemicals/ emissions from oil and gas by analyzing cancer incidences near oil fields in the Amazon basin of Ecuador (Sebastian et al. 2001). The basis for this study is to trace the effect of untreated waste products leaked or deposited, into water sources for drinking and household use. To analyze the data, they make use of a comparative statistical approach, that is, by comparing an expected number of cancer cases over a 10 year period, from 1989 - 1998 (3.5 cases for males and 4 for females) to observed ones (8 cases for males and 2 for females). Their results reveal an effect of 10 additional cancer patients within the area and conclude that the observed prevalence of cancer among people who live closer to the oil fields could be as a result of greater exposure to the oil contaminants than those who live at a greater distance.

Most studies recording health effects of oil and gas development are carried out after a major spill. A study of residents of an oil spill site (the Tasman Spirit oil spill) in Karachi, Pakistan revealed similar results. The study employed the use of two control groups and a main study group selected with respect to the distance from the incidence site. From their study they find that symptoms decrease with increased distance from the spill site suggesting that there is a higher degree of exposure as residents live closer to the spill sites. The following were reported as symptoms experienced by residents within closer proximity to the oil spills: increased incidences of headaches, nausea, vomiting and respiratory difficulty closer proximity to the spill site (Janjua, et al. 2006). Respiratory difficulty in particular is attributed to a higher level of exposure to pollutants (Sebastian and Hurting, 2004).

Janjua et al. (2006) identify the acute health effects of the Tasman oil spill in Pakistan through the use of a logistic regression model, they modelled one treatment and two control groups relative to distance from the oil spill site. They find that the more exposed group reports more cases of health illnesses (90%) associated with the spill compared with the control groups (70 and 85%) suggesting that these illnesses are most likely a result of exposure to the oil spill.

Other studies on the effects of oil pollution suggest the occurrence of future risks associated with exposure rather than effects that are feasible within a short time from the spill hence the true impact may not be estimated properly if only short term analysis were conducted. Grigalunas et al. (1986) reported no short term effects of the Amoco Cadiz spill on human health at the time of their study. They also suggest the possibility of long term effects. Such future health risks may occur not from direct exposure, but secondary (or indirect) exposure. For example, Solomon and Jensen (2010) state that "trace amounts of harmful metals can accumulate over time in fish tissues potentially increasing future health hazards from consumption of large fin fish such as tuna and mackerel".

A study of the Sea Empress oil spill reveals other potential health effects involving psychological and physical symptoms. Lyons et al. (1999) examined 539 exposed residents to the spill and compared this with 550 unexposed residents all randomly selected, to see if residents living closer to the oil spill sites reported more cases of physical and psychological health problems. They do this by administering surveys that include questions on demographics, a checklist of symptoms and questions that address the respondents' beliefs about the possible health effects from oil activities. Although the author allowed for potential bias in the response of the sample population due to other factors such as unemployment, anxiety and fear from other issues, it was still observed that the exposed study group were still significantly worse than the control groups (Lyons, et al.1999).

Epidemiological results from the study of the Prestige oil spill show that the risk of lower respiratory tract infections increase with exposure in terms of length in days and hours. The researchers use questionnaires to acquire quantitative and qualitative information on symptoms of respiratory illnesses. Data were collected from 6780 fishermen association members which included about 63% of workers that could have been exposed from clean-up operations. The authors find that among clean-up workers, more symptoms of lower respiratory tract illnesses were reported. They also made sure to exclude respondents that report anxiety to remove any bias in reporting. An interesting result that they find is that the risk of lower respiratory tract illnesses increased as the hours and days of exposure increased but after more time elapsed, this risk starts to decrease (Zock et al. 2007).

Another similar epidemiological study on the Prestige spill was carried out using cleanup volunteers, manual and “high-pressured” workers; they also included a control group. The researchers find that endocrine toxicity is potentially present due to “xenobiotics” in the oil that result in endocrine disruptions; this therefore poses risks to those exposed and requires protective strategies to avoid those health risks (Pérez-Cadahía et al, 2007).

2.3 Health Effects of Air Pollution

For studies that look at air pollution exposures, air samples are usually collected and using similar procedures the sample are tested and compared to existing air quality requirements as a means of identifying degrees of exposure. Health outcomes of people within the area are then analyzed to see if the increased air pollutants result in increased incidences of illness. McKenzie et al. (2012) use EPA guidance to analyze cancer and non-cancer risks in Garfield County, Colorado for a human health impact assessment. Their major finding was that residents living within 0.5 mile from a well had a higher risk than those living further away. Specifically, meaning that there is a chance for 10 cancer cases out of a million people within 0.5 miles which is greater than the 6 cases in a million chance for those living within a greater distance from a well (above 0.5 miles).

One major aspect missing from most of the literature on the health effects of air pollution, and particularly oil and gas development, on health, is to provide economic interpretations to these results. That is, the answer to the question 'What are the economic/monetary costs to society of increased oil and gas development?' is theretofore unknown.

There have been various applications of estimating health costs in other scenarios, particularly in the field of health economics. Estimating health costs is vital for making appropriate policy making decisions. Estimating health costs from pollution requires careful analysis and caution as there are other external variables that interact and could also be potential sources of illnesses. As regards air and water pollution as sources of poor health outcomes from oil and gas development, I will be looking specifically at air pollution.

Asthma has been found to be one of the resulting environmentally attributable illnesses (Ladrigan et al., 2002; Witter et al., 2008; Trasande and Liu 2008; Minnesota Center for Environmental Advocacy, 2006; Collaborative for Health and Environment; 2005). Children are found to be more susceptible to environmentally-related asthma because their respiratory organs are not yet fully developed (Sly and Flack, 2008; Neidell, 2004). This is what informs my use of children as a population of interest in this study. In Garfield Colorado, a study by the Saccomano

Institute, revealed that children in this region had an increased asthma rate, although exact values were not reported in this paper (Witter et al., 2008, 26).

There are various studies on the effects of air pollution on child health, particularly asthma (Neidell, 2004). These studies have shown the negative outcomes of poor air quality on children. One such finding has been carried out through the use of test score data. The rationale behind this is that air pollution results in poor health outcomes like asthma (Currie and Neidell, 2005) and poor health leads to child absenteeism, reduction in concentration levels and cognitive ability (Currie et. al, 2009); thus test scores have been used as an indicator for measuring effects on children. Ham et al. (2011) also uses this approach in a study of children in California using a fixed effects quantile regression model. The pollution variables included in this model are similar to that of Neidell (2004), only including the $PM_{2.5}$ pollutants. Results show that a reduction in air pollution leads to "increased academic performance by a small but significant amount for both math and English scores". This has implications for examining welfare effects on children because if academic performance influences their future earnings (Ham et al. 2011), reductions in pollution levels will have both immediate and future benefits.

Neidell (2004) uses a fixed effects regression model with data on asthma hospitalizations of children using zip codes in California and grouped into age brackets of 0-1, 1-3, 3-6, 6-12 and 12-18. He regresses these against data on pollutants in the area (CO, O_3, PM_{10}, and NO_2) as well as control variables including the children's demographics and temperature variables of the study area. His major findings are that the carbon monoxide results in increased asthma hospitalizations and that pollutants in general do not seem to influence infants as much as they do older children.

Including demographic information in this type of analysis is important especially for welfare analysis as it is necessary to see if a certain group of people are more at risk due to their demographic characteristics. For example, are less well to do households more prone to living closer to these sources of pollution? Morales et al. (2002) states that the socio-economic characteristics of people are a huge determinant of their health outcomes.

One major gap in the literature is the lack of information on the economic implications of these established effects. That is, what are the individual and societal costs of these health effects? Depending on the results of this analysis, it will be interesting to be able to estimate and see what the economic costs are from oil and gas development as they affect the population.

More general studies have estimated health costs of environmentally attributable asthma (Minnesota Center for Environmental Advocacy, 2006; Collaborative for Health and Environment; 2005; Trasande and Liu; 2008); costs from mortality and costs from morbidity.

2.4 Estimating Costs of Health Outcomes

Two frameworks are generally used in calculating morbidity costs: the cost of illness method (COI) as well as stated/revealed preference measures (willingness to pay, willingness to accept techniques and production functions). WTP and WTA measures are calculated using hedonic models or contingent valuation methods (Rozan, 2001; Hall et al, 2010). The COI and contingent valuation methods are usually used separately, though Rozan (2001) proposes the use of both simultaneously under the notion that using just COI frameworks tends to isolate private costs that are captured through the use of contingent valuation methods.

Hedonic models are able to capture factors such as the "avoidance strategy" (e.g., Neidell 2004). Neidell (2004) indicates families with information on the levels of pollution in an area tend to move away from the area to avoid risks associated with exposure. Another important aspect of avoidance behavior, also related to the socio-economic characteristics of households, is that it is possible that families of higher socio-economic backgrounds have better chances and resources that enable them to exhibit this behavior whilst families of lower socio-economic backgrounds may not have the resources needed to move and thus are the ones that are largely affected since they probably live closer to these sources of risk. This avoidance strategy manifests itself through housing prices and can be deduced through hedonic methods.

Bennett (2013) used hedonic methods to look at the impact of hydraulic fracturing on house prices in Weld County, Colorado. She finds that there was no negative effect of fracking on housing prices in the area and this is because people in the area seem to be benefitting a lot from the influx of employment activities into the area and this somehow overshadows whatever concerns that might be present over hydraulic fracturing activities.

Natalieva et al. (2005) use the health production function to estimate the costs of air emissions from oil extraction in Kazakhstan. The author focuses on the direct costs through losses in income and leisure time. From the model used, the direct costs are a function of the symptoms experienced by the residents. The authors look at the costs experienced by residents of two different regions, one region with a higher level of pollution than the other. They find that the health costs for the region exposed to oil extraction was higher than that of the unexposed region.

The cost-of-illness framework, which I am mostly interested in, calculates the direct and indirect costs associated with the illness in question (Hall et al. 2010). It is important to know that there are three methods of calculating direct costs: the top-down method, bottom-up method and the incremental/regression/econometric method.

Direct costs include medical costs from inpatient, emergency room, outpatient care, medications, while indirect costs are costs from work days lost, or school days lost in this context, and premature death (Rozan, 2001). Table VIII in the appendix shows a summary of all cost-of illness studies that are reviewed subsequently including information on various sources of data.

Cost of illness studies have a wide variety of data sources and the choices of which to use usually depend on the goals of the analysis (such as whether the study is prevalence based or incidence based and also on the purpose of the study). This is discussed in more detail in the subsequent paragraphs.

Weiss et al. (1992) using the bottom-up method in calculating direct costs, estimated costs of illness attributable to asthma in the United States and find that this was about $11 billion (2015 inflation-adjusted dollars) in 1990, with inpatient costs taking up the largest category of expenses at about $2.8 billion (2015 inflation-adjusted). They also find that indirect costs from school absences was the largest (35.7%) of the indirect costs, suggesting that children were strongly affected among others in the population.

In a 1997 study, Smith et al. (1997) use the 1987 National Medical Expenditure Survey (NMES) to calculate the economic costs of asthma in the United States. They use the cost of illness technique and find the total costs of asthma to be $8.4 billion (2015 inflation-adjusted), with the direct costs accounting for about 88.1% of the total costs and the indirect costs accounting for about $979 million (2015 inflation-adjusted).

The top down method of calculating direct costs usually involves the use of a population attributed fraction. This fraction measures the “proportion of a disease that can be attributed to the disease or to a risk factor” thus environmental factors have been developed to account for environmental health risks through the environmentally attributable fraction (EAF) (Segel, 2006). Ladrigan et al. (2002) used this approach in their study, producing an EAF that has been applied by many other researchers (e.g., Schuler et al. 2006; Davies, 2005).

One limitation of this approach is its inability to account for confounding factors that influence the prevalence of a disease. The bottom-up approach on the other hand looks at and calculates costs as a whole. Direct costs of the illness are calculated by finding the average costs of each treatment attributed to that disease and then multiplying it by the disease prevalence. The disease prevalence in this case is the amount or degree of use of the service/treatment (Segel, 2006).

Lastly, the econometric/incremental approach looks at the differences between a population with the disease and another without the disease (kind of like a treatment and control group analysis). These two groups usually have similar confounding factors but differ in just the presence of the disease of interest. There are also two ways of doing the econometric approach: the mean differences approach and the multi-stage regression approach. The mean-difference looks at the mean costs of the two groups –one with the disease and one without and the aim of this is to identify incremental differences between the two. This type of approach is useful for providing cost estimates at an individual level. It provides the cost per case estimate rather than a total cost estimate (Segel, 2006). The second type of incremental cost approach is used to check for the likelihood of requiring medical care and the excess costs of that care if received (Segel, 2006). Selection of the type of approach depends on the overall goals of the study and data requirements. For example, if a cost per case estimation is required the best approach may be the regression based approach.

Lozano et al. (1999) look specifically at childhood asthma and find that health care costs per child with asthma was $1583.76 (2015 inflation-adjusted) per year, compared with $468 per child per year without asthma. This study however does not provide specific results of direct and indirect costs. Using the National Medical Expenditure Survey (NMES) data, the authors were able to compute per capita costs of asthma from information of costs required for treating a child affected by asthma. The costs are reflected through payments for ambulatory care, inpatient stays, emergency department visits and costs of prescribed asthma drugs. They also took into account the source of payments, tracking whether payments were out-of pocket or from full or part insurance coverage. In calculating the per child cost of asthma, they calculate the total health care expenditure from the data and then deduce the percentage of children that have asthma and with that find how much they spend per year on asthma expenses. They compare these costs per child per year of asthma to costs per year per child of those without asthma.

Wang et al. (2005) state the importance of estimating costs on a per capita basis in their study of the direct and indirect costs of asthma among school-aged children in the United States. They state that “although national estimates provide policy makers with information on the economic effects of asthma in the Unites States, information on the per capita costs of asthma among children with asthma helps to determine the benefits of a successful asthma intervention”.

Their results show that direct costs of medical services and treatments for children with asthma was $597.30 (2015 inflation-adjusted) per child in 1996 and indirect costs were 2.48 days in number of missed school days per child. Other costs included, loss of productivity of parents, and life time earnings lost through loss of future productivity as a result of death from asthma. Generally, according to their results the total economic costs of asthma in 1996 was $1178.22 (2015 inflation-adjusted). This approach is similar to that of Lozano et al. (1999) and Barnett and

Nurmagambetov (2010) discussed earlier, because they all analyze and present results on a per capita basis. The only difference is in the kinds of models that each paper uses, Wang et al (2005) on one hand use a linear regression model while Barnett and Nurmagambetov (2010) use a GLS model with Lozano et al.(1999) deviating from the regression-model approach.

In a 2010 study, two-part models (Generalized linear models) were used to estimate the incremental direct and indirect costs of asthma in the US from 2002-2007, which were also calculated on a per capita basis. Results showed that the total incremental direct costs of asthma in 2007 was about $63.1 billion (2015 inflation-adjusted) and indirect costs of morbidity (loss in work days and school days) was about $4.1 billion (2015 inflation-adjusted) in 2009 dollars (Barnett and Nurmagambetov, 2010). They find the direct incremental costs of asthma per person per year to be $3,673.40 (2015 inflation-adjusted), incremental indirect costs as $339.27 (2015 inflation-adjusted) for workers and $104.83 (2015 inflation-adjusted) for students with respect to loss in work and school days.

Here, I use the incremental cost estimation approach as I am interested to see how much of an effect and what costs that an additional well has on health outcomes for the population of interest. Also because it is regression-based, it will be possible to model other independent and control variables specific to this study.

Hasegawa (2013) look at child asthma hospitalizations in the US from 2000-2009; they look specifically at direct costs of the illness. Their findings show that the total direct costs of child asthma within that time period was estimated to be $502 million annually with younger children between the ages of 0-4 accounting for the larger part of the cost. However, they state that this increase was because hospitalizations became more costly over the years.

In calculating asthma costs attributed to environmental factors, the cost of illness method is used to account for the environmental factor since asthma is attributed to many elements including environmental pollution. Researchers have tackled this by developing models to account for the contribution of environmental factors to illnesses. Landrigan et al. (2002) calculate the costs of environmentally attributable asthma among American children using an "environmentally attributable fraction" (EAF) model; they find that the total annual costs of asthma as at the time of the study was about $2.6 billion dollars (2015 inflation-adjusted).

The total cost of illnesses from environmental factors is calculated as:

"Disease rate * EAF * Population size * Cost per case"

Where Cost per case refers to the discounted lifetime expenses of an environmentally attributable illness (Ladrigan et al. 2002) i.e. the direct and indirect costs of the illness while the population size and disease rate is determined by either the "incidence" or "prevalence" of the disease. Incidence-based costs are calculated based on all expenses from the onset of the disease while prevalence-based costs just estimate costs in a given period of time, for example, the annual costs of that disease (Tarricone, 2006; Pervin et al. 2008; Segel, 2006). Most studies use the prevalence based cost approach because it is easier in the sense that it doesn't require as much data as incidence based costs (Pervin, 2008). Ladrigan et al. (2002) use the prevalence of the disease because morbidity can be reduced by abatement of the elements of exposure.

The EPA uses the incidence based method in a slightly different context. In looking at diseases such as asthma, they try to estimate the benefits of reduced pollution from revised standards of National Ambient Air Quality Standards (NAAQS). This is done through a Regulatory Impact Analysis (RIA) that provides the public with information on the costs and benefits of attaining the newly revised standards. They combine WTP and COI methods to come up with estimates for valuing benefits from reduction in ambient air pollution. WTP estimates showed the value people had on reducing future risks of illness from air pollution. Their COI estimates on the other hand showed costs expended in treating or avoiding adverse health effects. COI results show that the central estimate of the value per statistical incidence for asthma admissions for people between ages 0-64 is $16,000 and the value for hospital emergency department visits is $430. Value per statistical incidence here refers to "the value of small changes in risk of an effect for a large number of people". That is, it is the value from having avoided the risk of an effect. The central estimate of $16,000 for asthma admissions reflects the costs (from the Agency for Healthcare Research and Quality (AHRQ)) that the people incur from mean costs of hospital care, average length of hospital stay and for the weighted values of other asthma related illnesses. An important point to note about the estimates for asthma admissions by the EPA is that they assume that air-pollution related hospital admissions are unscheduled and also agree that this might understate benefits of air-pollution reductions if they were on the contrary scheduled admissions.

The central estimate of $430 for emergency department visits are derived from mean COI values reported by two studies: Smith et al. (1997) and Stanford et al. (1999).

WTP provided results for asthma exacerbations and not for newly developed asthma cases which was valued at $45 per incidence. This estimate was derived from a study by Rowe and Chestnut, 1986 who calculated WTP estimates of avoiding a "bad asthma day" defined by the patients themselves; they also define asthma exacerbation as "a day of moderate to worse asthma" (United States Environmental Protection Agency, 2012).

One challenge of the EAF model discussed earlier, is the process of arriving at a value for the Environmentally Attributable Fraction. One way researchers have been able to overcome this is through the use of the "DELPHI" technique. This technique involves the use of expert panels who consult a vast amount of literature and jointly come up with a consensus fraction estimate. The disadvantage of this technique is the fact that it is time consuming and might be costly giving the amount of resources that would be needed. Ladrigan et al. (2002) come up with an EAF with a range of 10-35% as the percentage of environmentally attributable illnesses under the basis that environmental factors affect child asthma symptoms 30% of the time. This EAF fraction has been used in many other studies (Minnesota Center for Environmental Advocacy, 2006; Collaborative for Health and Environment; 2005; Trasande and Liu; 2008).

Trasande and Liu (2008) use the same methods by applying an environmentally attributable fraction of the same range of 10-35% to calculating costs of environmental illnesses in American children such as asthma, lead poisoning, etc. Their estimates in 2008 show that costs of childhood asthma from environmental related factors was about $2.3 billion implying that there was an increase of $0.3 billion dollars within the time period of 2002 and 2008. Not to infer conclusively about the implications of the results; results however suggests that policy making decisions were not able to address the issues of childhood exposure (Trasande and Liu, 2008) or could also imply that costs of medical care increased within that time period. In support of my latter assumption; Hasegawa et al. (2013) state that between the years of 2000 and 2009 "mechanical ventilation use and hospital charges for asthma increased significantly in the United States."

One major data source for calculating health costs of asthma and the link to air pollution is asthma hospitalizations data (Trasande and Liu 2008; Ladrigan et al.2002; Neidel 2004). This type of data provides information on the direct and indirect costs of asthma needed for the health cost calculations. This source of data is particularly useful for my study because it also provides information on the addresses/ locations of the patients which will be needed for establishing causal linkages between locations of oil wells and patients' residences through a proximity analysis.

2.5 GIS Applications

The overall analysis will require data on respiratory hospitalizations and Geographical Information Systems (GIS) data on well locations. GIS data provides information on specific locations of the population of interest- in this case children often through a process of "address geocoding". Nuckols et al. (2004) say that "applications of GIS includes locating the study population by geocoding addresses (assigning mapping coordinates), using proximity analysis of contaminant source as a surrogate for exposure, and integrating environmental monitoring data into the analysis of the health outcomes." This demonstrates the potential of using GIS data in this analysis.

Several epidemiological studies on health outcomes relative to the environment use GIS for their analysis. Maantay (2007) uses GIS in looking at asthma and air pollution in the Bronx, New York. Methods involved matching up data on asthma hospitalizations using the addresses of the patients. These GIS methods aid in proximity analysis for establishing causal relationships between asthma outcomes and air pollution in the area.

According to Vine et al. (1997) "GIS technology is an important tool for environmental health researchers as it can be used to support or suggest hypotheses regarding disease causation through the conduct of relatively quick and inexpensive ecologic studies using existing databases and easily computerized data". This is in line with the objectives and circumstances of my study as using GIS for this study provides a quicker and less expensive means of approaching this study given time and monetary limitations.

CHAPTER 3: MODEL SPECIFICATION

3.1 Methods

As this paper is focused on identifying the human health effects of oil and gas production and possibly estimating costs of such effects. The steps towards achieving the overall objective will be to:

- Establish a relationship between oil and gas development and child respiratory illness hospitalizations in Colorado
- Estimate costs of child respiratory hospitalizations in Colorado

In trying to establish a relationship between oil and gas development and child respiratory illness hospitalizations in the state, a difference-in-difference approach is used. This method is used to measure causal effects of a certain incidence, program by capturing before and after effects usually between two time periods (Lechner, 2011). Because I assume that the treatment effect (increased oil and gas development) occurred within these two time periods: 2004 and 2013, the difference-in-difference approach helps to identify the mean effect of moving from one time period to another.

Although there are various difference-in difference strategies that researchers have used to estimate causal effects (Waldinger, 2014), the one for this paper follows Card and Krueger (1994). Which is simply:

Outcome = (Pre-Treatment – Post- Treatment) – (Pre-Control – Post- Control)

To reiterate, health costs are difficult to measure because most times their effects manifest over a long period of time. For policy making, information on the economic impact of oil and gas development needs to be available. This would require strong and unbiased results that are able to find a strong causal relationship between these negative effects and oil and gas development.

It would be interesting to apply the cost-of-illness framework in estimating the economic costs of respiratory hospitalizations from exposure to oil and gas activities as most studies from the literature focus on general outdoor air pollution and few look at oil and gas in particular. But this is entirely dependent upon findings from the analysis.

Another application of existing and reviewed literature is by incorporating GIS techniques first of all by matching up health data and spatial data through zip code information, secondly, GIS to extract information on the total number of wells located in each zip code which is useful for intended regression analysis and thirdly, by creating map visualizations of the data in identifying the relationships between oil and gas activities and child respiratory hospitalization cases in the state.

Generally, using GIS as a technique is usually applied to epidemiological studies with a gap in applying this same method to identifying and estimating economic effects. I would argue that great potentials lie in using GIS for economic studies and could provide breakthroughs in terms of meeting data limitations and should thus be given more consideration.

Therefore, the overall model for my econometric analysis incorporating ideas from literature will require respiratory hospitalization data for children in Colorado as well as GIS data on residential location relative to the oil wells and other important control variables of socio-economic characteristics of children to be studied. According to Morales et al. (2002), “evidence suggests that social and economic factors are important determinants of health”. Also, in the North Carolina Institute of Medicine publication, it was stated that a person’s race and ethnicity along with their income, educational achievement and other determinants are part of the key predictions of health status. They also conclude that people with fewer years of education, lower incomes, less accumulated wealth, living in poor neighborhoods/substandard housing have worse health outcomes.

Another major contribution of this study is the use of children as a population of interest. One reviewed literature that looks at the health effects of natural gas operations in Pennsylvania looks at infant health (Hill, 2012) so this study provides another perspective to these type of studies, providing more information for decision/policy makers in the state.

3.2 Variables of Interest

The general empirical model which is informed by methods used in previous studies already stated (McKenzie et al.2012; Hill 2013; Neidell 2004; Ham et al. 2011) will include the variables:

Table I : Description of Variables

Variable	Description	Units	Data Source
Dependent Variables			
Cases of Illnesses/1000 people	The number of illnesses in each zip per 1000 people in the area	Count/1000	AHRQ-HCUP-SID 2004 and 2013/ US Census 2000 and 2010
Respiratory Illness Count	The count of illness cases in each zip code	Count	AHRQ-HCUP-SID 2004 and 2013
Independent Variables			
Number of Producing Wells	Well count of producing wells in each zip code	Count	COGCC 2004 and 2013
Wells/Square-Mile	The number of wells per square-mile of land area in each zip code area	Count/Square-Mile	COGCC 2004 and 2013/ US Census TIGERLINE 2000 and 2010
	The median household income in each zip code area		
Median Household Income	The squared value of median household income in each zip	Thousands	US Census 2000 and 2000
Median Income – Squared	The total number of people in each zip code area		US Census 2000 and 2010

Total Population	The percent rural population from the total population in each zip code area	Count	US Census 2000 and 2010
Percent Rural Population		%	US Census 2000 and 2010
Dummies			
Dummy- Wells in both Years	DV; = 1 if there are wells in both years	1/0	COGCC 2004 and 2013
Dummy- Wells in 2004	DV; = 1 if there are wells in just 2004	1/0	COGCC 2004 and 2013
Dummy- Wells in 2013	DV; = 1 if there are wells in just 2013	1/0	COGCC 2004 and 2013

3.3. Multiple regression model specification

Multiple regression analysis will be used to estimate the health impact of oil and gas production in the area. The dependent variable in the model is the health impact measured by respiratory inpatient cases for children in Colorado; most available studies have used other indicators such as infant health (Hill 2013, Schwartz, 2004). Their reason being that, “Although criticisms have arisen because of the lack of depth of the IMR (Infant Mortality Rate) as a measure of population health, the IMR has remained an important indicator of health for whole populations suggesting that factors that affect the health of entire populations have an impact on the mortality rate of infants” (Reidpath & Allothey, 2003). I do not use child mortality rates here because it will be difficult to infer from such data as only a minute number actually die from respiratory illnesses. Rather, I concentrate on estimating the morbidity effects through the use of data on child respiratory inpatient cases in the state. My approach of using respiratory hospitalizations data therefore builds upon existing literature and contributes another perspective to establishing causal relationships between oil and gas operations and human health.

Residential locations using zip code information will be compared with the number of licensed drilling sites (number of wells & wells/square-mile) in each zip code. This is to see if living closer to oil drilling sites increases the risk of poor health outcomes from respiratory illnesses. Similar studies have also used this approach of proximity (Janjua et al. 2006; Hill 2013).

Other control variables include: Socio-economic characteristics of the population in each zip code (Income, rural-urban population). The reason for the inclusion of socio-economic factors in the regression as control variables is in line with other studies that show that there is a relationship between socio-economic status and health (Neidell 2004; Ham et al. 2011).

I include models using three dummies to represent the presence or lack of for wells in either 2004/2013 or both years. The first independent dummy variable takes the value of 1 if there are wells in both years and 0 if there are no wells existing in both 2004 and 2013. For the second dummy, 1 is turned on if there are wells in 2004 and 0 if there are no wells in 2004 and the third dummy takes the value of 1 if there are wells in 2013 and 0 if there are no wells in 2013.

The model used for the OLS regression analysis are specified below:

3.3.1 Baseline Model

$$\text{Cases of Illnesses/ 1000 People}_{2004\text{-}2013} = \alpha_i + \beta_1\text{Wells/SquareMile}_{2004\text{-}2013} + \beta_2\text{MedianIncome}_{2004\text{-}2013} + \beta_3\text{MedianIncome\^{}2}_{2004\text{-}2013} + \beta_4\text{PercentRuralPopulation}_{2004\text{-}2013}$$

3.4. Data Analysis

I run a baseline model with variables already mentioned above as well as other supplementary models that use "number of respiratory illnesses in each zip code" as opposed to the cases of respiratory illnesses per 1000 people in each zip code area. The choice of cases by 1000 people in the population is because of the small number of cases reported in the data. The key independent variable of the baseline model is the "Number of Wells per Square-mile". In other supplemental models, I use the original count of oil and gas wells in each zip code. Other independent variables included are: Median Household Income, total population and percent rural population. The decision to choose wells per square-mile was as a result of visually observing the distribution of wells in the state. Zip codes normalize for population but vary by land area, and I felt it would be important to account for the variability in population density relative to the concentration of oil and gas wells located in the areas.

I expect that my main independent variable from the baseline model (wells/square-mile) should be positive, because I expect that as more development occurs in terms of increased oil and gas activities from wells, more cases of respiratory illnesses should be observed, since the main hypothesis is that oil and gas operations reduce air quality and this in turn increases chances of respiratory illnesses among populations within close proximity to these activities in the area.

For the other independent variables included in the models, I include them to account for confounding factors that influence health outcomes and this is consistent with other literature (Rapaport and Bonthapally, 2012; McKenzie et al. 2014). I expect the coefficient on Median Income to be negative, so that as income increases, the cases of illnesses decrease. This could be because higher income people avoid areas of pollution (Neidell 2004) and can afford to move to better places of fewer to none oil and gas development areas. Another possible assumption is that people of higher income will tend to have better lifestyles and thus reduced cases of illnesses but if air pollution is the underlying cause of respiratory illness cases from oil and gas activities, this might not hold.

To test the significance of each variable, I evaluate each variable by conducting t-tests and evaluating results at the 10%, 5% and 1% levels of statistical significance. The overall model is tested for significance using F-tests. The data are analyzed using STATA and GRETL statistical software.

3.5 Hypothesis

Ho: Oil and gas development has no significant effect on Child Health Outcomes in Colorado

$H_{1:}$ Oil and gas wells has a significant effect on the Child Health Outcomes in Colorado

CHAPTER 4: DATA

4.1. Oil and Gas Data

Data on respiratory illnesses among children ages 19 and below in Colorado were acquired from the Agency for Health Care Research and Quality (AHRQ), Health Care Cost and Utilization Project (HCUP), State Inpatient Database (SID). The State Inpatient Database of the AHRQ contains information on inpatient stays from states that participate in the H-CUP. It captures about 97 percent of US community hospital discharges including patients who are insured and those who are uninsured. For different states, government data organizations provide data for all community hospitals but if data is provided by a private data organization, they only report data for member hospital. The SID database also provides a cost-to-charge ratio that enables researchers to McKenzie (2014) calculate the cost of healthcare for various illnesses. Data for years 2004 and 2013 were used for the study to capture the changes in health outcomes following two different levels of oil and gas development, the latter year representing increased activities. This is represented visually in Figure I. Figure II on the other hand shows the number of producing wells for both years in each zip code. From Figure II we can also see that while many zip codes do not have any wells, some few zip codes have very many wells. In 2013 for example, some zip codes had as much as 3000 wells. The figure also tells us that from 2004 until 2013 the number of zip codes with a high number of wells increased.

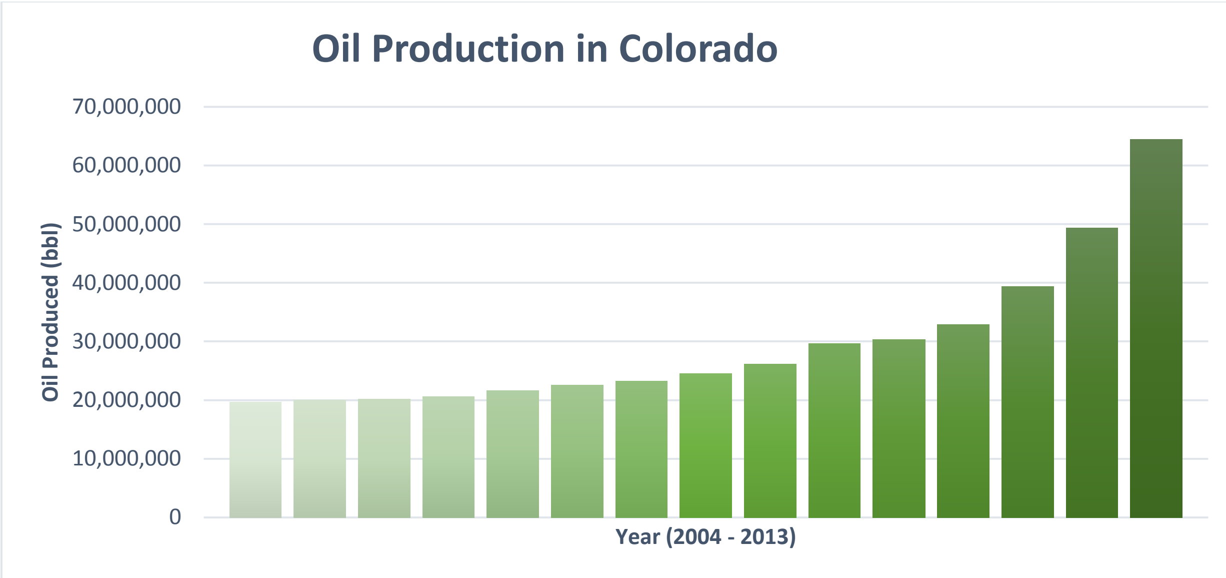

Figure I: Shows the graphical representation of the progression of oil and gas development over the years (COGCC)

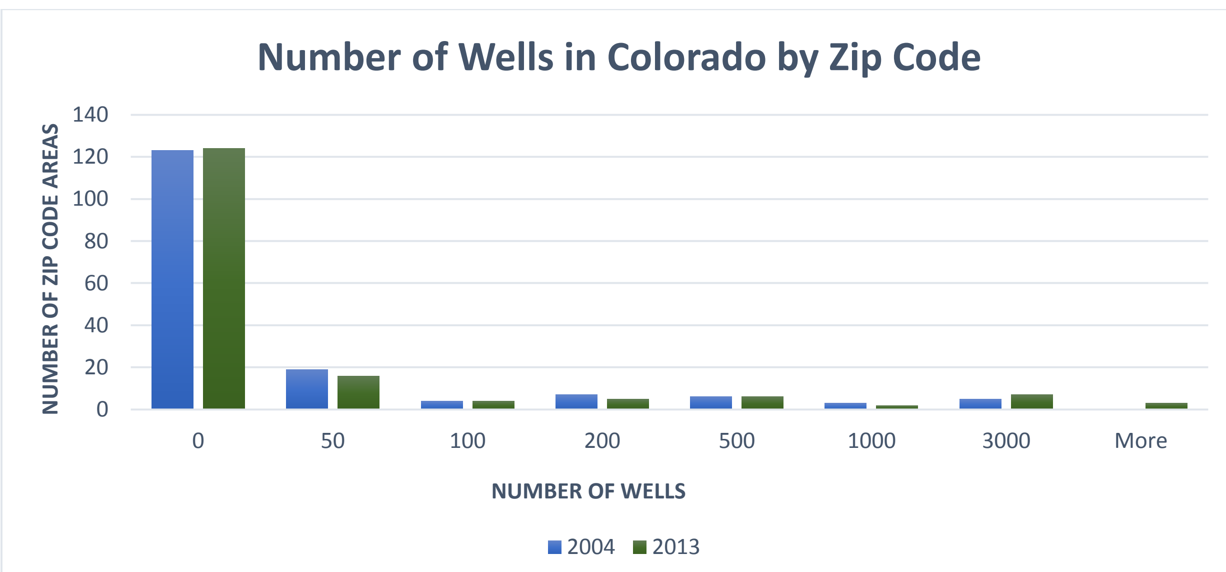

Figure II: Graphical representation of the total number of wells in Colorado for years 2004 and 2013

4.2. Health Data

An extraction process was required to obtain the information of interest from the large database of the AHRQ-HCUP-SID. The database was filtered based on age (0 – 19) and based on children with respiratory illnesses. Illnesses were reported using ICD-9-CM codes (International Statistical Classification of Diseases, 9th edition, Clinical Modification); these codes represent diagnosis for various health encounters. Children with respiratory illnesses were selected using codes 460- 519.9. For sensitivity analysis, data on "Fractures" were also extracted from the entire Colorado SID database using codes 820.00- 826.1.

Initially there were 551 observations for respiratory illnesses in 2004 and 539 in 2013 for children between ages 0-19 (observations are zip code areas). However, one problem identified in the data was that a lot of observations were missing information on inpatient visits for respiratory illnesses. In order to be able to account for true zeros and non-zeros in the data, I filter the entire SID data set including all diseases reported and select only zip codes that report 10 cases and above (this is purely based on intuition) within each year. I assume that in a given year there should be at least 10 illnesses reported for each zip code and thus any zip code that reports below 10 cases is excluded. With this I assume that any missing observation of respiratory illness is actually a true zero. This exclusion and filtering process further reduces the number of observations used for the analysis to 167 zip codes for both years also making sure that the observations for both years match up in order to be able to measure the inter-temporal impact between the two years. Summary statistics in Table I shows a greater number of respiratory cases recorded for year 2013. Filtering the data does not change the results by much because most of the zip codes that were dropped did not have wells in them.

4.3. Demographic Data

Demographic data from the 2000 Census and 2010 Census were used to provide information on other independent variables of interest such as Median Household Income and Rural-Urban location. These variables account for other factors that might affect the health outcomes of the population of interest. The results of the summary statistics in Table I reveal that the average median income between years 2004 and 2014 increases by about 30.7%.

Data on the number of wells in the state of Colorado for the years 2004 and 2013 were obtained from the Colorado Oil and Gas Commission (COGCC) website. This database provides geographic information (latitude and longitude) on the locations of producing wells (oil and gas) in the state of Colorado. This information was used to identify the counts of wells in each zip code using the ArcGIS software. For the baseline model of which wells/square-mile was used, this was calculated using the TIGERLINE shape file data (2000 and 2010) from the U.S. Census Bureau. Summary statistics in Table I show that there was an increase in the number of wells in 2013 from 2004. In 2004 the average number of wells out of all 167 zip codes measured was about 86 while in 2013 this number increase by almost half to about 168 wells.

CHAPTER 5: RESULTS

5.1 GIS Map Results

Some intuition about relationships going on in the data can be deduced from the GIS map layouts. In Figures III and V representing 2004 data, wells seem to be clustered around lower income areas and these areas look to have higher cases of respiratory illnesses for the most part. Also, most high income areas had no wells in them but high income areas that did have wells clustered around them tended to also have higher cases of respiratory illnesses. Some of these high income areas with many wells and high respiratory illness cases are the Broomfield and Weld county areas. Visually, this might suggest that there could be a positive relationship between the wells and respiratory illness cases as we would expect that since it is a high income area, people should have more improved health outcomes.

For 2013, in Figures VII and VIII the same trends as that of 2004 can be seen. It however seems like there is substantial variation across zip codes, because in some low income zip code areas, there are no wells but respiratory cases are still high. One reason for this could be that since people there have low incomes they may be overall more susceptible to poorer health outcomes. For some other lower zip code areas there are a few number of wells and also a few cases of respiratory illnesses. Another interesting thing that can be seen is that a few high income zip codes have high cases of respiratory illnesses and no wells. I would expect that better income and no wells will have lower cases of respiratory illnesses if there was a correlation between the two. But one explanation for this could be as a result of the size of the zip and the number of people in the zip code area which I already tried to account for by looking at the data in terms of cases by population and running the regression analysis in terms of the number of wells per square-mile.

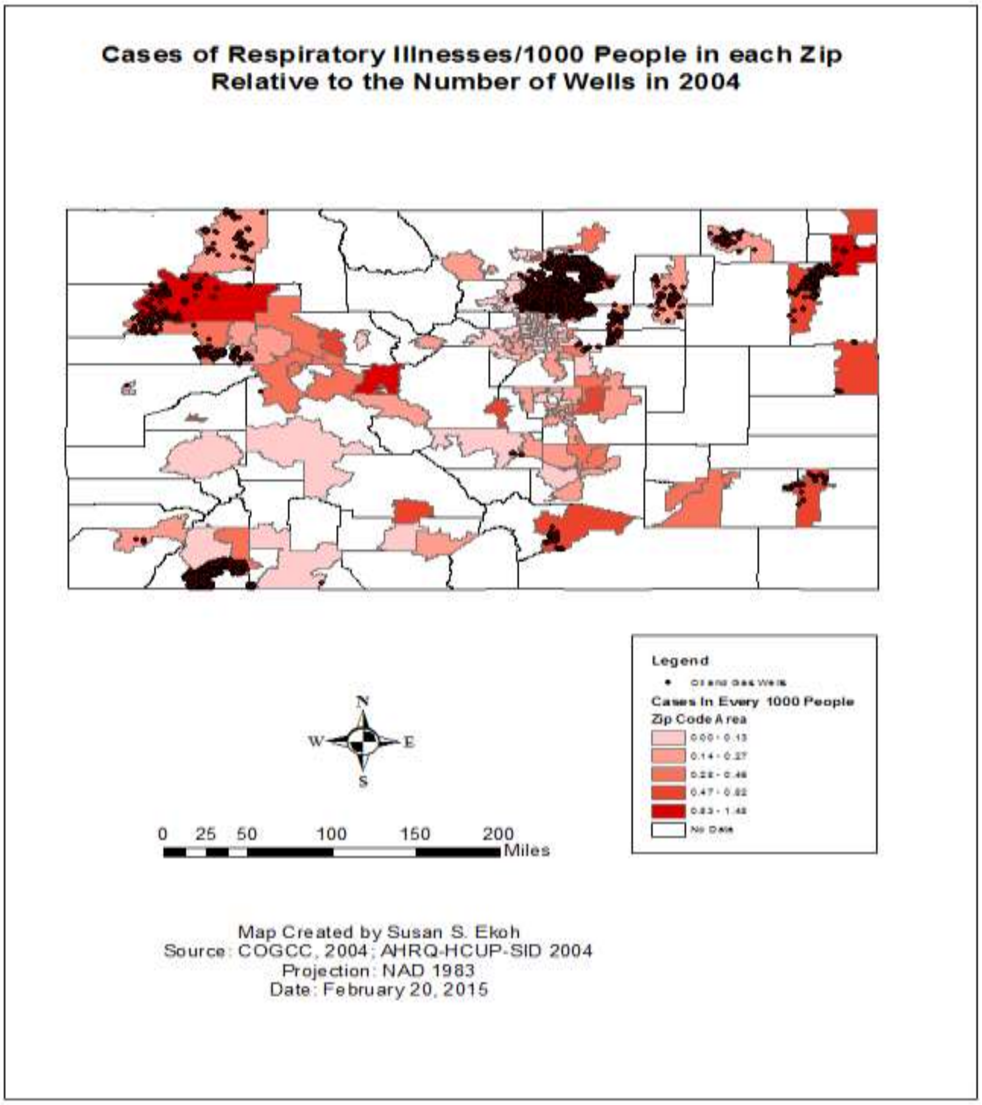

Figure III: Cases of Respiratory Illnesses/1000 People in each Zip Relative to the Number of Wells in 2004

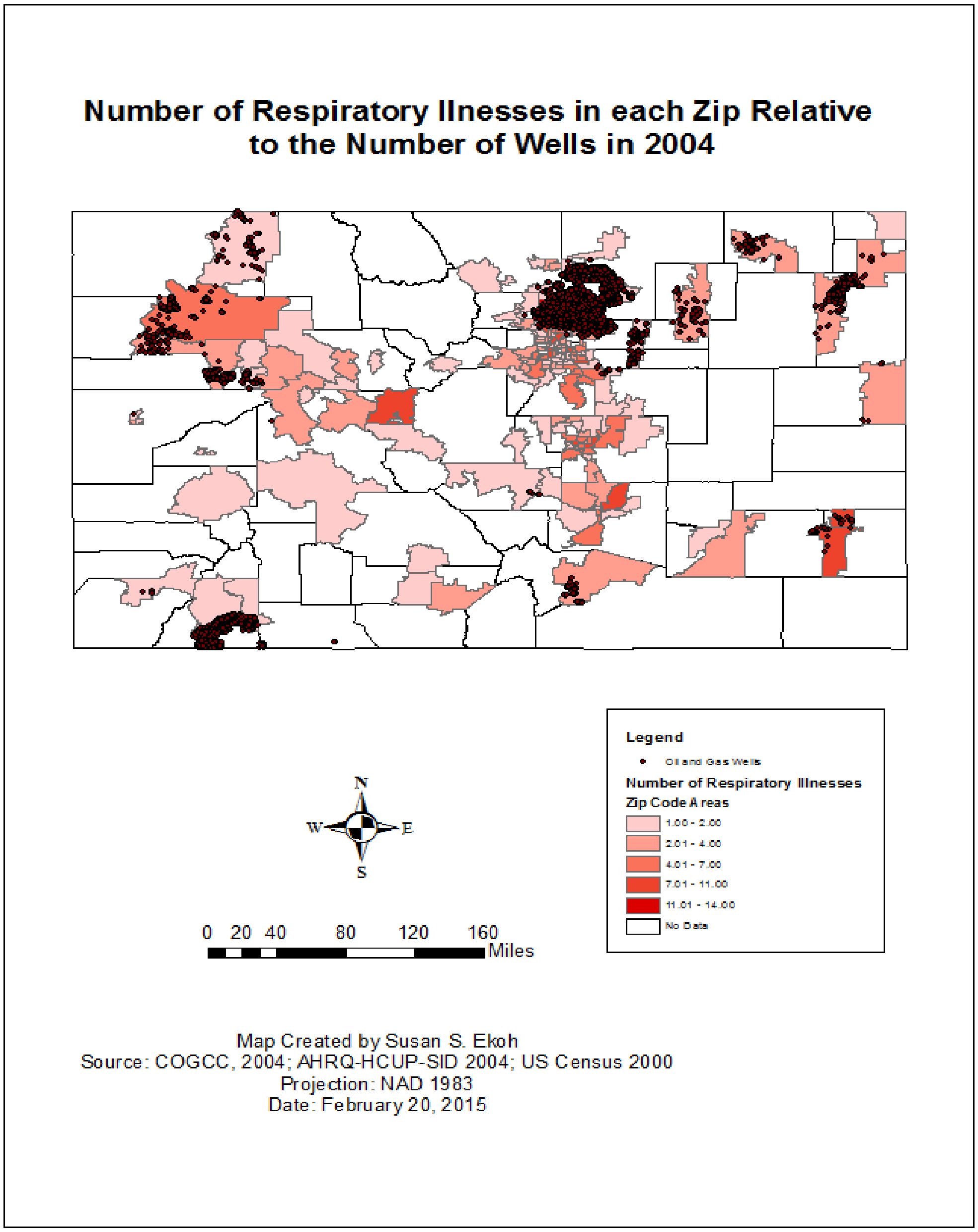

Figure IV: Number of Respiratory Illnesses in each Zip Relative to the Number of Wells in 2004

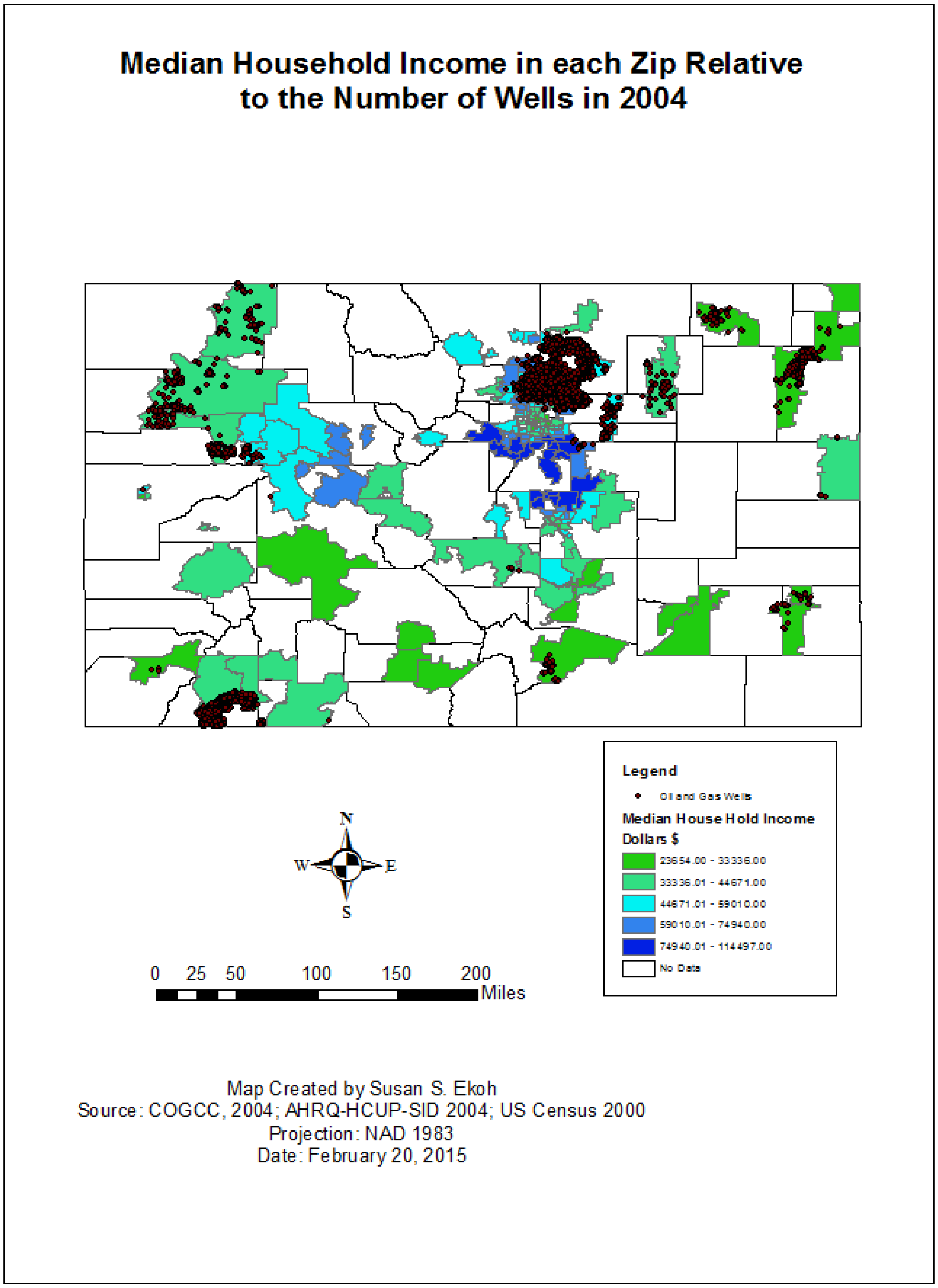

Figure V: Median Household Income in each Zip Relative to the Number of Wells in 2004

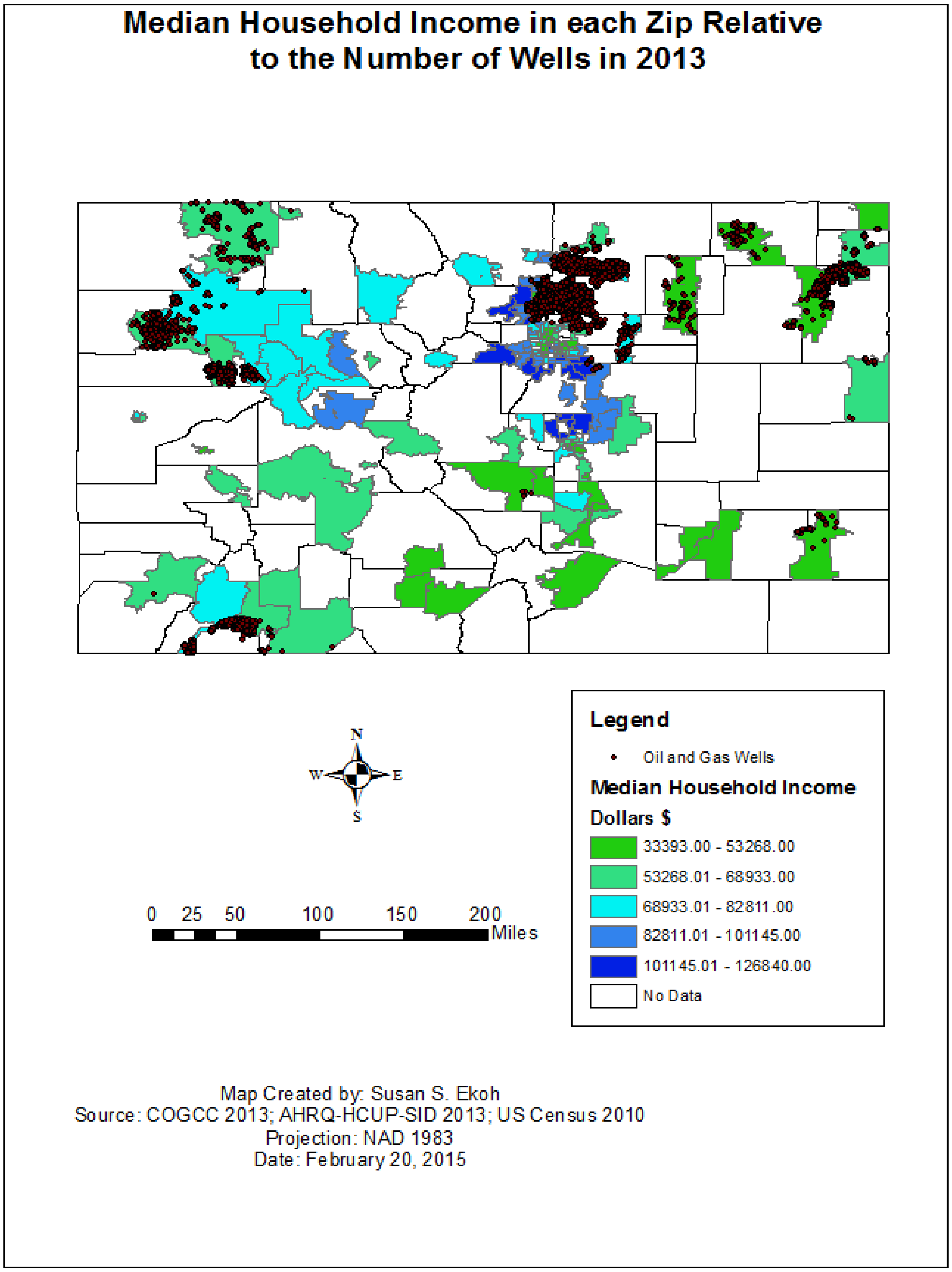

Figure VI: Median Household Income in each Zip Relative to the Number of Wells in 2013

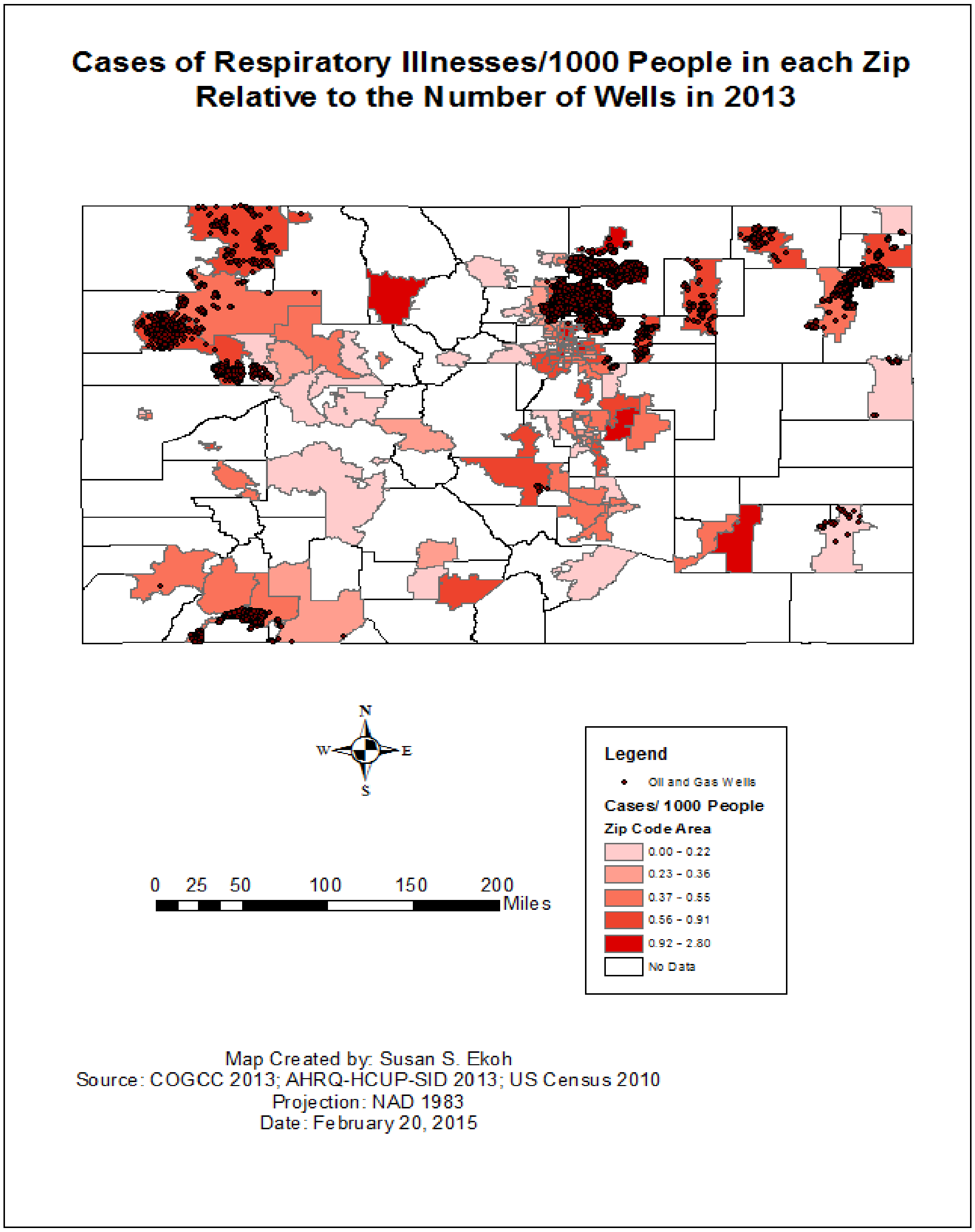

Figure VII: Cases of Respiratory Illnesses/1000 People in each Zip Relative to the Number of Wells in 2013

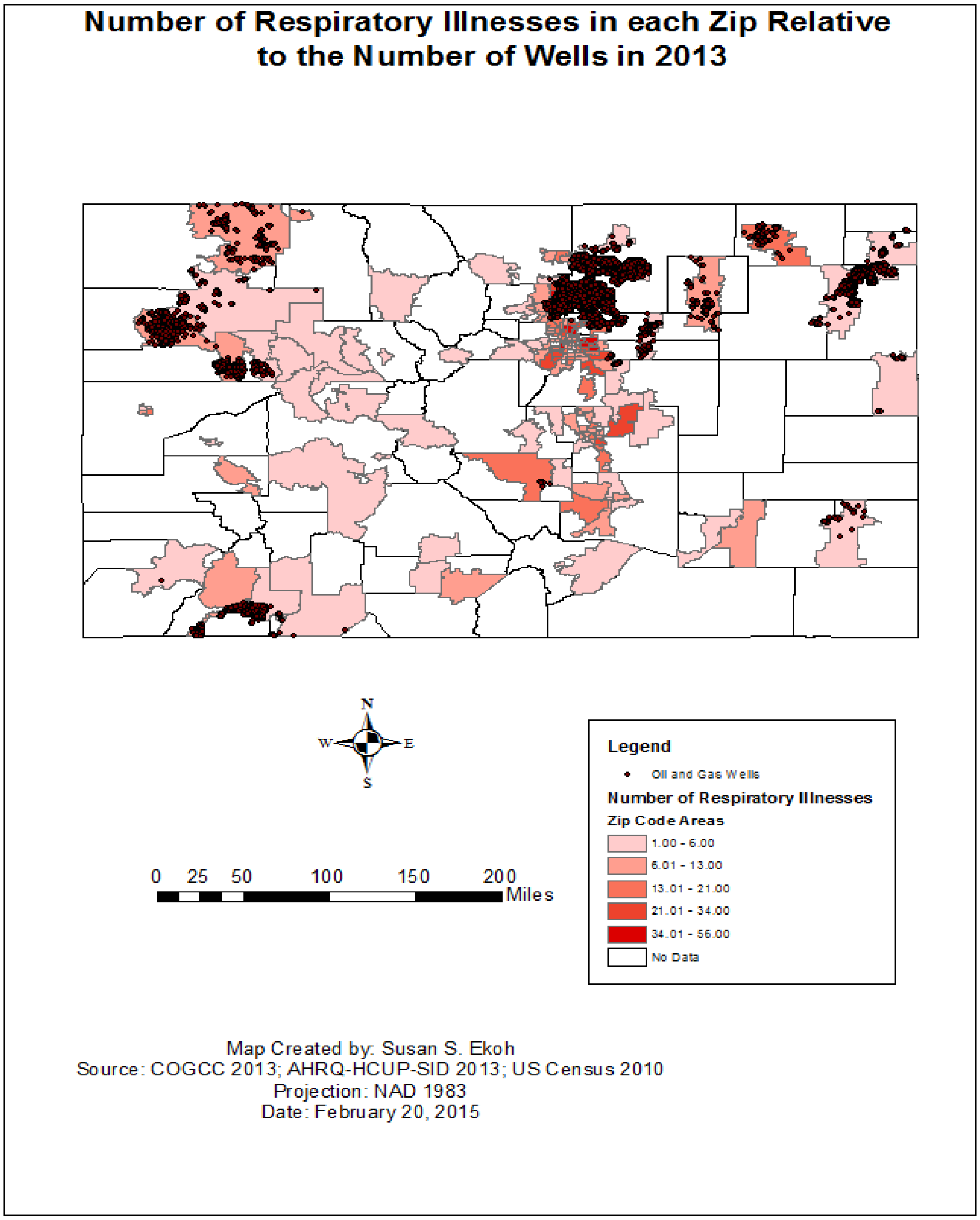

Figure VIII: Number of Respiratory Illnesses in each Zip Relative to the Number of Wells in 2013

5.2 Regression Results

The difference-in-difference approach is used to test for causality and to account for the possibility of inter-temporal effects of oil and gas development with the hypothesis that observed health outcomes should be more visible when looked at over time as oil and gas activities increased over time.

Baseline Model:

$$\text{Cases of Illnesses/ 1000 People}_{2004\text{-}2013} = \alpha_i + \beta_1\text{Wells/SquareMile}_{2004\text{-}2013} + \beta_2\text{MedianIncome}_{2004\text{-}2013} + \beta_3\text{MedianIncomeSquared}_{2004\text{-}2013} + \beta_4\text{PercentRuralPopulation}_{2004\text{-}2010}$$

In Tables III, IV, V, VI, VII and VIII, I present individual year results for 2004, 2013 and the difference in difference results for respiratory illnesses and compare them with results from sensitivity analysis (control) where I look at inpatient fracture cases in the state.

My baseline model with results in Table III and IV for respiratory illnesses and fractures respectively, includes Cases of Illnesses per 1000 people in the area as the dependent variable and Wells/Square-mile as the main independent variable of interest. Other independent variables included are Median Income and Percent Rural Population.

The choice of fractures for a sensitivity analysis is to provide stronger inference on the causal relationship between oil and gas development and respiratory illnesses with the major assumption that oil and gas development has no relationship between fractures (in terms of air pollution). So I expect that the coefficients on the Wells variable for the respiratory illness model to be positive and that of fractures to be zero.

Other models, which can be found in the appendix, involve looking at the association of oil and gas on illnesses using different measures; using number of wells in each zip code (Table VI and VIII), using the count of illnesses from inpatient visits (Tables VI, VII and IX), using Cases/1000 people (Tables III, IV, VIII and X) and using dummies representing the presence of wells in either 2004, 2013 or in both years (Tables IX and X). These supplementary models are just there for additional robustness testing. Unlike the baseline model that accounts for the variability in land area and population density.

Tables III and IV shows results of the baseline model; regression results here represent the impact of oil and gas development through number of wells/ square-mile in each zip code area of Colorado on inpatient cases of illnesses among every thousand people in the area. Parameter estimates, their standard errors and levels of significance are all reported in Tables III and IV.

Looking at my main independent variable of interest-wells/square-mile, results in 2013 show a coefficient estimate of 0.0087, which is significant at the 5% level. This means that as the number of wells per square-mile increases by one unit, the number of cases of respiratory illnesses in the area among every 1000 people increases by 0.0087, implying that in that year, oil and gas development was correlated with incidences of respiratory inpatient cases in the state (note that only 167 zip codes were observed). However, the coefficient becomes insignificant when tested with robust standard errors. For 2004 on the other hand, results show a positive expected impact, but results are insignificant at all levels.

The coefficient on the difference-in-differences estimate is -0.079 and significant at the 10% level but also becomes highly significant at the 1% level when the robust-standard error is used. The coefficient estimate just implies that as the number of wells per square mile in a zip code increased over time from 2004 until 2013 by one unit, the number of cases in every 1000 people decreased by 0.079. Signifying that a negative relationship existed with oil and gas development over time (from 2004 until 2013) with the cases of respiratory illnesses recorded. It could also be that over time other factors that influence respiratory illnesses such as exposure to air emissions from vehicles interact with the possible effect of oil and gas development and this makes it difficult to establish a causal relationship between oil and gas and respiratory health solely. The negative effect of oil and gas on respiratory illnesses could be explained by the percent rural population variable. The variable increased by a little over time, its coefficient -0.0089 which is significant at the 10% level suggests that as more people moved into rural areas, respiratory illnesses decreased. This could mean that the cases of illnesses reported were more in urban areas than rural areas.

Comparing the above results with a sensitivity analysis using fracture inpatient cases as a control group, I find that negative and significant relationships exist between oil and gas and fractures in both years 2004 and 2013 with the latter being significant at the 10% level and highly significant at the 1% level when the robust standard error is used. The coefficient on the independent variable- wells per square mile is -0.016, implying that a one unit increase in the number of wells per square mile results in a decrease of 0.016 for reported fracture cases per 1000 people in the population. The similarity of the results for respiratory and fracture illnesses and their possible effect from oil and gas, is that both are significant only in 2013 and not in 2004. The difference however is that the treatment has a positive relationship with oil and gas development and the control has the opposite, a negative relationship. Also, the difference-in-difference results for fractures as a control is negative, but not significant.

Overall, results show that over time, oil and gas development had a negative relationship with respiratory illnesses and had no relationship with fractures. And since there was a positive relationship in 2013, it could suggest that the possible effects of increased oil and gas activities exhibits a short term effect and does not quite hold over time. Another explanation could be that if oil and gas activities does impact health outcomes over a long period of time, maybe these effects are reflected in other types of illnesses that this particular data does not capture.

The lack of significance of the wells per square mile variable over time does not allow for a cost of illness estimation from the potential effect of oil and gas development.

Looking at the variable Median Income, over time, the cases of respiratory illnesses decrease by 0.0144 as the median income increases by a $1000. This coefficient is statistically significant at the 1% level. The implications of this result possibly means that over time, as people become better off economically which is often reflected in improved incomes, health outcomes are also improved.

Comparing these results for the baseline model with the other supplementary models in the appendix previously mentioned, the results do not vary much across both when looking at the significant variables. The significant variables in all models are consistent with respect to the signs of the coefficients. The supplementary models are not preferable to the baseline model because they do not take into consideration jointly, the variability in land density and population density.

5.2.1. Summary Statistics

Table II: Summary Statistics for Respiratory Illness Analysis

Variable	Mean	Minimum	Maximum	Standard Deviation	Number of Observations
Wells/Square	1.61	0.00	118.22	9.79	167
Mile	1.77	0.00	37.29	6.27	167
Number of	86.01	0.00	2482.0	327.73	167
Wells	168.29	0.00	4505	630.02	167
Cases/	0.23	0.00	1.48	0.22	167
Population(‘000)	0.52	0.00	2.80	0.34	167
Cases/	0.16	0.00	1.18	0.17	167
Population(‘000)	0.08	0.00	0.41	0.08	167

Fractures					
Number of Illnesses (Respiratory)	3.39 11.93	0.000 0.000	14.00 56.00	2.57 11.03	167 167
Number of Illnesses (Fractures)	2.89 1.86	0.00 0.00	14.00 11.00	2.65 2.00	167 167
Median Income	49800.65 71900.72	23654 33393	114497 126840	16756.08 22377.48	167 167
Total Population	21089.89 22913.67	1349 1498	68492 69588	14600.47 14237.83	167 167
Percent Rural Population	9.72 9.06	0.000 0.000	61.09 65.08	14.42 15.06	167 167
Note: 2004 summary statistics are above while 2013 are below for each variable					

5.2.2. Regression Results for Baseline Model

Table III: Regressions on the Association of Oil and Gas on Respiratory Illnesses

Variable	2004 Respiratory Illness	2013 Respiratory Illness	Inter-temporal Difference (Respiratory Illness)
Constant	0.3787* (0.1295) (0.0996)*	0.9055* (0.2702) (0.2674)	-0.4358* (0.0596) (0.0653)*
Wells/Square -Mile	0.0008 (0.0016) (0.0008)	0.0087** (0.0042) (0.0053)*	-0.0079** (0.0039) (0.0022)*
Median Income	-0.0065 (0.0048) (0.0033)**	-0.0083 (0.0075) (0.0073)	-0.0144* (0.0059) (0.0060)*
Income Squared	0.00004 (0.0000) (0.0000)	0.0000 (0.0000) (0.0000)	0.0001** (0.0000) (0.0000)***
Percent Rural Population	0.0068* (0.0011) (0.0017)*	-0.0006 (0.0018) (0.0019)	-0.0089** (0.0037) (0.0072)
R-Squared	0.2286	0.0645	0.1015
Number of Observations	**167**	167	167

Note: Standard errors are in parenthesis; Statistical significance is tested at the *1%, **5% and ***10% levels. The robust standard error for each coefficient is included in parentheses below the traditional standard error.

Table IV: Regressions on the Association of Oil and Gas on Fractures (Control)			
Variable	2004 Fractures	2013 Fractures	Inter-temporal Difference (Fractures)
Constant	0.2636** (0.1096) (0.1193)**	0.1659* (0.0615) (0.0669)*	0.0389 (0.0302) (0.0267)
Wells/Square -Mile	-0.0007 (0.0008) (0.0003)*	-0.0016*** (0.0009) (0.0006)*	-0.0008 (0.0019) (0.0007)
Median Income	-0.0049 (0.0041) (0.0045)	-0.0021 (0.0017) (0.0018)	-0.0063** (0.0029) (0.0033)***
Income Squared	0.0000 (0.0000) (0.0000)	0.0000 (0.0000) (0.0000)	0.0000*** (0.0000) (0.0000)**
Percent Rural Population	0.0022** (0.0004) (0.0006)	0.0003 (0.0004) (0.0006)	0.0057* (0.0019) (0.0026)**
R-Squared	0.0499	0.0375	0.0724
Number of Observations	167	167	167
Note: Standard errors are in parenthesis; Statistical significance is tested at the *1%, **5% and ***10% levels			

From the results of the baseline model discussed above, I believe that the results reveal the possibility respiratory illnesses being more prevalent in urban areas than rural areas and thus allowing for the potential causal effects of other sources of air pollution. I run another model to see whether the presence of wells differs across rural and urban areas. For this model, I introduce an interaction variable of wells per square mile and the percent rural population. Regression results are in table V below:

Table V: Regressions on the Association of Oil and Gas on Respiratory Illnesses with Interaction			
Variable	2004 Respiratory Illness	2013 Respiratory Illness	Inter-temporal Difference (Respiratory Illness)
Constant	0.3669* (0.1301) (0.0983)*	0.9040* (0.2713) (0.2666)*	-0.4341* (0.0600) (0.0669)*

Wells/Square -Mile	0.0013 (0.0017) (0.0003)*	0.0091 (0.0042) (0.0049)***	-0.0070 (0.0048) (0.0049)
Median Income	-0.0061 (0.0048) (0.0033)***	-0.0083 (0.0075) (0.0073)	-0.0143** (0.0059) (0.0061)**
Income Squared	0.00004 (0.00004) (0.00003)	0.00003 (0.00004) (0.00005)	0.0001*** (0.0000) (0.0000)***
Percent Rural Population	0.0071* (0.0011) (0.0018)*	-0.0006 (0.0018) (0.0019)	-0.0087** (0.0038) (0.0075)
Wells/Square -Mile * Percent Rural Population	-0.0002 (0.0002) (0.0001)**	-0.00003 (0.0003) (0.0004)	-0.0032 (0.0011) (0.0011)
R-Squared	0.2332	0.0646	0.1020
Number of Observations	167	167	167
Note: Standard errors are in parenthesis; Statistical significance is tested at the *1%, **5% and ***10% levels			

The results of this model support my findings as the coefficient for the interaction between wells per square mile and the percent rural population is not significant over time and for the most part also applies to the individual year results. The lack of significance of the interaction variable suggests that the cases of respiratory cases reported in the data are not strongly related to whether wells exist in rural or urban zip codes.

CHAPTER 6: CONCLUSION

Results from the baseline model of this study shows that oil and gas could potentially have a short-term correlation with respiratory health outcomes of children in Colorado particularly as seen in results for 2013, but over time may not have an effect on respiratory outcomes.

One possible reason why results for 2004 do not show significance could be as a result of the small sample of cases observed in the data for that year. It could also be that since 2004 had fewer wells and thus not as much oil and gas activities and also possibly lower density of people around wells, respiratory health outcomes were not affected.

Results also show that the cases of respiratory illnesses seen in the data occurred more among the urban population than the rural population, suggesting that the respiratory illnesses could be as a result of exposure to other sources of air pollution and that’s why over time the effect of oil and gas is negatively related with the respiratory cases analyzed.

Also worth noting are the results represented in the GIS maps created in Figures III, V, VI and VII. They showed that wells tend to be concentrated more in areas where the number of cases per 1000 population are the highest and close to the highest. Areas with lower cases per 1000 people have little to no number of cases among every 1000 people. But these are all unique to the zip codes, as some do not follow this trend which is discussed further in the results section. Regarding income as seen in Figures V and VI which represent 2004 and 2013 respectively, lower to middle income areas have more wells concentrated in and around them. Suggesting that higher income members of the population will tend to live further away from oil and gas wells if they think it might be harmful. They will also be able to make this move out of such areas because economically they will have the means to do so.

These results are important in determining what groups of the population are affected by the location of oil and gas wells and the potential effects of living close to these activities.

Finally, cost estimates were not calculated as I do not find significant positive relationships between oil and gas development and respiratory illnesses over time.

6.1 Limitations of the Study and Further Research Suggestions

One limitation of the study is obviously the limited number of observations only 167 zip codes were analyzed out of about 482 zip code areas in Colorado for 2004 and 525 for 2013. This could be that either the data did not report some respiratory illness cases or that children with these illnesses in the state of Colorado received other kinds of health services that were not inpatient services, or that they received inpatient health care services outside of the state. The latter being more unlikely. Witter et al. (2008) mention this in their paper, they state that since not all asthma cases in particular are recorded in inpatient data other types of data (emergency room and outpatient data) might be of better use.

Also in establishing more significant causal relationships between respiratory health outcomes and oil and gas development, it would have been better to have data on the smoking habits of the child's parents to account for that potential confounding factor.

Since state inpatient data were collected for each case with location of patient recorded at the zip code level I was not able to calculate the particular distance of patients from the wells. This also would have been a better measure for establishing the causal impact of air pollution through oil and gas activities. Finally, calculating the costs of respiratory illnesses are dependent on the ability of the models to predict significant causal relationships which wasn't the case in this paper. So in future research, if positive relationships are found to be true for the effect of oil and gas development on respiratory health over time, the total costs can be deduced by multiplying the number of predicted cases by the cost per case.

REFERENCES

Allen, D. T., Torres, V. M., Thomas, J., Sullivan, D. W., Harrison, M., Hendler, A., & Seinfeld, J. H. (2013). Measurements of methane emissions at natural gas production sites in the United States. *Proceedings of the National Academy of Sciences*, *110*(44), 17768-17773.

Barnett A.G., Williams G.M., Schwartz J., Neller A.H., Best T.L., and Petroeschevsky A.L., and Simpson R.W. (2005). Air Pollution and Child Respiratory Health: A Case Cross-Over Study in Australia and New Zealand. *American Journal of Respiratory Critical Care Med* Vol 171. Pp 1271-1278, 2005

Bennett A., (2013). Impact of Hydraulic Fracturing on Housing Values in Weld County, Colorado: A Hedonic Analysis. Master's Thesis Colorado State University.

Card D., and Krueger A.B. (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania, *The American Economic Review,* 84, 772-793

Clark, C. R., Walter, M. K., Ferguson, P. W., & Katchen, M. (1988). Comparative dermal carcinogenesis of shale and petroleum-derived distillates. *Toxicology and Industrial Health*, *4*(1), 11-22.

Colborn, T., Kwiatkowski, C., Schultz, K., & Bachran, M. (2011). Natural gas operations from a public health perspective. *Human and Ecological Risk Assessment: An International Journal*, *17*(5), 1039-1056.

Currie J., Neidell M. and Schmieder F., (2009). Air Pollution and Infant Health: Lessons from New Jersey. *Journal of Health Economics* 28 (2009) 688-703.

Doyle, J. (1994). Crude Awakening: The Oil Mess in America: Wasting Energy, Jobs & the Environment. Friends of the Earth.Energy Intell. Group, (2003). Polity Score and Reserves Table. http://www.energyuintel.com/grounds/tables.htm

Environmental Performance Index. (2013) www.epiyale.edu/our-methods/health-effects

Epstein P.R, Selber J, (2002). Oil: A Life Cycle Analysis of Its Health and Environmental Effects. *Boston: Center Health Glob. Environ*, Har. Med Sch.

Energy from Shale (2013) http://www.energyfromshale.org/fracking-benefits

Ferrar, K. J., Michanowicz, D. R., Christen, C. L., Mulcahy, N., Malone, S. L., & Sharma, R. K. (2013). Assessment of effluent contaminants from three facilities discharging Marcellus Shale wastewater to surface waters in Pennsylvania. *Environmental science & technology*, *47*(7), 3472-3481.

Finewood M.H, Stroup L.J, (2012). Fracking and the Neoliberalization of the Hydro-Social Cycle in Pennsylvania's Marcellus Shale. *Journal of Contemporary Water Research and Education* Vol. 147, Issue 1, 72-79.

Grigalunas T.A, Anderson R.C, Brown G.M, Congar R, Meade N.F, Sorenson P.E, (1986). Estimating the Cost of Oil Spills: Lessons from the *Amaco Cadiz* Incident. *Marine Resource Economics, Vol 2(3).*

Grimmer, G., Dettbarn, G., Brune, H., Deutsch-Wenzel, R., & Misfeld, J. (1982). Quantification of the carcinogenic effect of polycyclic aromatic hydrocarbons in used engine oil by topical application onto the skin of mice. *International archives of occupational and environmental health*, *50*(1), 95-100.

Hall J.V., Brajer V. and Lurmann F.W., (2010). Air Pollution, Health and Economic Benefits-Lessons from 20 years of Analysis. Ecological Economics 69 (2010) 2590-2597

Ham J.C, Zweig J.S, Avol E., (2011) Pollution Test Scores and the Distribution of Academic Achievement: Evidence from California Schools 2002-2008 (working paper)

Hill E.L, 2013. Shale Gas Development and Infant Health: Evidence from Pennsylvania. Working, Paper. The Charles H. Dyson School of Applied Economics & Management, Cornell University, Ithaca, New York 14853-7801 USA.

Horwitt D, (2009). Drilling Around the Law. Environ Working Group. Washington D.C.

International Agency for Research on Cancer (1989). Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Man: Occupational Exposures to Petroleum Refining; Crude Oil and Major Petroleum Fuels.

Janjua N.Z, Kasi P.M, Nawaz H, Farooqii S.Z, Khuwaja U.B, Hassan N, Jafri S.N, Lufti S.A, Kadir M.M, Sathiakumar N, (2006). Acute Health *BMC Public Health 2006: 6:84* doi: 10.1186/1471-2458-6-84

Landrigan P.J, Schechter C.B., Lipton J.M, Fahs C. and Schwartz J. (2002) Environmental Pollutants and Disease in American Children: Estimates of Morbidity, Mortality, and Costs for Lead Poisoning, Asthma, Cancer, and Developmental Disabilities. *Environmental Health Perspectives* 110: 721-728

Lauver L.S (2012). Environmental Health Advocacy: An Overview of Natural Gas Drilling in Northeast Pennsylvania and Implications for Pediatric Nursing. *Journal of Pedriatic Nursing* Vol.27 Issue 4: 383-389.

Lechner, M. (2011). The Estimation of Causal Effects by Difference-in-Difference Methods. Now.

Lyons R.A, Temple J.M.F, Evans D, Fone D.A, Palmer S.R, (1999). Acute Health Effects of the Sea Empress Oil Spill *J Epidemiol Community Health* 53:306-310. http://jech.bmj.com/content/53/5/306.full.pdf#page=1&view=FitH Accessed 7/12/2013

Maantay J. (2007). Asthma and Air Pollution in the Bronx: Methodological and Data Considerations in Using GIS for Environmental Justice and Health Research. Health and Place 13 (2007) 32-56

McKenzie, L.M, Witter, R.Z., Newman L.S., Adgate J.L., (2012) Human Health Risk Assessment of Air Emissions from Development of Unconventional Natural Gas Resources. *Science of the Total Environment*, doi:10.1016/j.scitotenv.2012.02.018

Meyer, B. R., Bazan, L. W., Jacot, R. H., & Lattibeaudiere, M. G. (2010). Optimization of multiple transverse hydraulic fractures in horizontal wellbores. In *SPE Unconventional Gas Conference*. Society of Petroleum Engine

Netalieva, I., Wesseler, J., & Heijman, W. (2005). Health costs caused by oil extraction air emissions and the benefits from abatement: the case of Kazakhstan. *Energy Policy*, *33*(9), 1169-1177.

Nukols J.R., Ward M.H., Jarup L., (2004). Using Geographic Information Systems for Exposure Assessment in Environmental Epidemiology Studies. Environmental Health Perspectives, Vol 112, No 9, 2004 pp. 1007-1015.

O'Rourke D and Connolly S, (2003). Just Oil? The Distribution of Environmental and Social Effects of Oil Production and Consumption. *Annu.Rev.Environ.Resour.2003*.28:587-617

Osborn, S. G., Vengosh, A., Warner, N. R., & Jackson, R. B. (2011). Methane contamination of drinking water accompanying gas-well drilling and hydraulic fracturing. *Proceedings of the National Academy of Sciences*, *108*(20), 8172-8176.

Ostro B. (2004). Outdoor Air Pollution: Assessing the Environmental Burden of Disease at National and Local Levels. Protection of the Human Environment, Geneva, World Health Organization (WHO Environmental Burden of Disease Series, No. 5).

Palinkas L.A, Peterson J.S, Russel J, Michael D.A, (1993).Community patterns of psychiatric disorders after the Exxon Valdez Oil Spill. *Am J. Psychiatry* 1993; 150:1517-1523

Pérez-Cadahía, B., Lafuente, A., Cabaleiro, T., Pásaro, E., Méndez, J., & Laffon, B. (2007). Initial study on the effects of Prestige oil on human health. *Environment international*, *33*(2), 176-185.

Ross M, (2001). Extraction Sectors and the Poor. Boston: Oxfam America.

Rozan, A. (2001). How to measure health costs induced by air pollution? *Revue Suisse d Economie Politique et de Statistique*, *137*(1), 103-116.

Rozell, D. J., & Reaven, S. J. (2012). Water pollution risk associated with natural gas extraction from the Marcellus Shale. *Risk Analysis*, *32*(8), 1382-1393.

Schmidt C.W, (2011). Blind Rush? Shale Gas Boom amid Human Health Questions. *Environ Health Perspect* Vol.119 (8):a348-a353

Schwartz J.D, (2013) Alternative Measures of Environmental Health Effects. The Metric. Yale &ES. www.epi.yale.edu/the-metric/alternative-measures-environmental-health-effects

Sebastian M.S, Armstrong B., Cordoba J.A, Stephens C, (2001). Exposures and Cancer Incidence near Oil Fields in the Amazon Basin of Ecuador. *Occup Environ Med 2001;* 58:517-522

Sebastian S.M and Hurting A.K, (2004). Oil Exploitation and Health in the Amazon Basin of Ecuador, Department of Public Health and Clinical Medicine, Umea University, Umea.

Smith K.R, (1993). Fuel Combustion, Air Pollution Exposure, and Health. The Situation in Developing Countries. *Annual Review of Energy and the Environment* Vol. 18: 529-566.

Soeder, D.J. 2010. The Marcellus Shale: Resources and reservations. Eos 91(32): 277.

Sly, P. D., & Flack, F. (2008). Susceptibility of children to environmental pollutants. *Annals of the New York Academy of Sciences*, *1140*(1), 163-183.

Solomon G.M, Janseen S, (2010). Health Effects of the Gulf Oil Spill. *JAMA* Vol. 304 No 10(Reprinted). http://jama.jamanetwork.com Accessed on 07/22/2013.

Tarricone R., (2006). Cost-of-Illness Analysis: What Room in Health Economics? Health Policy 77 (2006) 51-63

Trasande L and Liu Y. (2011). Reducing The Staggering Costs of Environmental Disease in Children, Estimated at $76.6 Billion In 2008. Health Affairs, 30, no.5 (2011): 863-870 (published online May 4, 2011; 10.1377/hlthaff.2010.1239.

United States Environmental Protection Agency. (2012). Regulatory Impact Analysis for the Final Revisions to the National Ambient Air Quality Standards for Particulate Matter.

Waldinger F. (2014) Lecture 3: Differences-in-Differences - University of Warwick. Retrieved from: http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0CB4QFjAA&url=http%3A%2F%2Fwww2.warwick.ac.uk%2Ffac%2Fsoc%2Feconomics%2Fstaff%2Fffwaldinger%2Fteaching%2Fec9a8%2Fslides%2Flecture_3_-

_did.pdf&ei=wOEBVdjaLoeYyQSPw4DABg&usg=AFQjCNGichArKweU0F_xcQxGaqFKfPVeLg&bvm=bv.87920726,d.aWw

Wilson J.S, Holland L.M, (1988). Periodic Response Difference in Mouse Epidermis Chronically Exposed to Crude-oils or BaP: males v females. *Toxicology* 50:83-94

Witter R., Stinson K., Sackett H., Putter S., Kinney G., Teitelbaum D., and Newman L., (2008). Potential Exposure-Related Human Effects of Oil and Gas Development: A White Paper

Zock J.P, Rodriguez-Trigo G, Pozo-Rodriguez F, Barbera J.A, Bouso L, Torralba Y, Anto J.M, Gomez F.P, Fuster C, Verea H.S, SEPAR-*Prestige* Study Group (2007). Prolonged Respiratory Symptoms in Clean-up Workers of the *Prestige* Oil Spill. *AM.J.Resp.Crit Care* 176:610-616

APPENDIX

Table VI: Regressions on the Association of Oil and Gas on Respiratory Illnesses Using Number of Wells as the Independent Variable and Number of Illnesses as the Dependent Variable

Variable	2004 Respiratory Illness	2004 Fractures	2013 Respiratory Illness	2013 Fractures	Inter-temporal Difference (Respiratory Illness)	Inter-temporal Difference (Fractures)
Constant	3.4712** (1.4402)	3.0632** (1.4551)	19.1093* (5.8474	2.0027 (1.2594)	-14.2995* (1.6339)	0.7122 (0.4453)
Number of Wells	-0.0000 (0.0005)	0.0002 (0.0005)	0.0012*** (0.0009)	-0.0001 (0.0002)	0.0014 (0.0022)	-0.0001 (0.0006)
Median Income	-0.0001 (0.0001)	-0.0001 (0.0001)	-0.0005* (0.0002)	-0.0001*** (0.0003)	-0.0006* (0.0002)	-0.0001** (0.0000)
Income Squared	4.12e-10 (4.67e-10)	6.04e-10 (4.72e-10)	2.34e-09** (1.02e-09)	2.86e-10 (2.19e-10)	2.18e-09** (9.43e-10)	5.34e-10** (2.57e-10)
Total Population	0.0001* (0.0000)	0.0001* (0.0000)	0.0006* (0.0000)	0.0001* (0.0097)	0.0001 (0.0001)	0.0001** (0.0000)
Percent Rural Population	-0.0047 (0.0066)	-0.0066 (0.0149)	-0.0004 (0.0449)	0.0098 (0.0097)	0.1296 (0.1043)	0.0244 (0.0284)
R-Squared	0.2947	0.3200	0.6045	0.4438	0.1406	0.0644
Number of Observations	167	167	167	167	167	167

Note: Standard errors are in parenthesis; Statistical significance is tested at the *1%, **5% and ***10% levels

I run the above model using the actual count of oil and gas wells in each zip code area as the key independent variable and also use the actual count of illnesses in each zip as the dependent variable. The reason why this does not cut it as the baseline model is because as discussed in the paper, land density is vital to this analysis since the zip codes vary by size and could impact the predictions if it isn't accounted for.

However, results from this model show that the number of wells over time has no significant effect on respiratory illnesses in the area.

Table VII: Regressions on the Association of Oil and Gas on Respiratory Illnesses Using "Wells/Square-Mile as Independent Variable and Number of Illnesses as Dependent Variable						
Variable	2004 Respiratory Illnesses	2004 Fractures	2013 Respiratory Illnesses	2013 Fractures	Inter-temporal Difference Respiratory Illnesses	Inter-temporal Difference Fractures
Constant	3.4956** (1.4399)	3.0773** (1.4544)	18.9032* (5.8554)	2.0215 (1.2598)	-15.2518* (1.6337)	0.7030*** (0.4447)
Wells/Square-Mile	-0.0082 (0.0177)	-0.0060 (0.0102)	0.0938 (0.0883)	0.0016 (0.0190)	0.0313 (0.1032)	-0.0962 (0.0281)
Median Income	-0.0006 (0.0005)	-0.0001 (0.0001)	-0.0005* (0.0002)	-0.0001*** (0.0000)	-0.0006* (0.0002)	-0.0001** (0.0000)
Income Squared	4.06e-10 (4.66e-10)	5.65e-10 (4.71e-10)	2.30e-09** (1.02e-09)	2.96e-10 (2.19e-10)	2.25e-09 (9.45e-10)**	5.41e-10** (2.57e-10)
Total Population	0.0001* (0.0000)	0.0001 (0.0000)	0.0006* (0.0449)	0.0001* (0.0000)	0.0001 (0.0001)	0.0001** (0.0000)
Percent Rural Population	-0.0054 (0.0147)	-0.0069 (0.0149)	0.0065 (0.0449)	0.0093 (0.0097)	-0.1292 (0.1045)	0.0238 (0.0284)
R-Squared	0.2956	0.3209	0.6031	0.4430	0.1389	0.0651
Number of Observations	167	167	167	167	167	167
Note: Standard errors are in parenthesis; Statistical significance is tested at the *1%, **5% and ***10% levels						

The above model looks at the effect of the number of wells per square- mile on the actual count of illnesses in the area. This model accounts for the land density factor. Results on the key independent variable of interest shows no significant relationship between oil and gas and respiratory illnesses. It also shows that median income had a negative relationship with illnesses for the year 2013 and also over time. This implies that better income results in better health outcomes.

Table VIII: Regressions on the Association of Oil and Gas on Respiratory Illnesses Using Cases/Population as Dependent Variable and Number of wells as Independent Variable						
Variable	2004 Respiratory Illness	2004 Fractures	2013 Respiratory Illness	2013 Fractures	Inter-temporal Difference (Respiratory Illness)	Inter-temporal Difference (Fractures)
Constant	0.3790* (0.1297)	0.2652** (0.1098)	-0.4384* (0.0637)	0.1638* (0.0612)	-0.4149* (0.0662)	0.0214 (0.0303)
Number of Wells	-0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0001)	-0.0000** (9.58e-06)	0.0000 (0.0001)	-0.0001*** (0.0000)
Median Income	-6.45e-06 (4.81e-06)	-5.11e-06 (4.08e-06)	-0.0000** (6.23e-06)	-2.07e-06 (1.70e-06)	-0.0000** (6.26e-06)	-7.76e-06** (3.07e-06)
Income Squared	3.57e-11 (4.19e-11)	4.65e-11 (3.54e-11)	6.69e-11*** (3.60e-11)	1.12e-11 (1.10e-11)	5.23e-11 (3.62e-11)	4.11e-11** (1.77e-11)
Percent Rural Population	0.0065* (0.0010)	0.0021** (0.0009)	0.0007 (0.0014)	0.0004 (0.0004)	0.0078** (0.0038)	0.0060* (0.0018)
R-Squared	0.2276	0.0477	0.0463	0.0458	0.0458	0.1021
Number of Observations	167	167	167	167	167	167
Note: Standard errors are in parenthesis; Statistical significance is tested at the *1%, **5% and ***10% levels						

The above model accounts for the variability in population density for zip code areas but also just looks at the effect of the number of wells for each zip code area. For the most part, the number of wells did not have an impact on respiratory illnesses in the area.

Table IX: Regression of Inter-Temporal Difference using Dummies for Number of Wells as Independent Variables and Number of Illnesses as Dependent Variable		
Variable	Respiratory Illness	Fractures
Constant	-14.4655* (1.6702)	0.8009*** (0.4542)
Dummy- Wells in both years	0.2072 (1.7935)	-0.4315 (0.4877)
Dummy- Wells in 2004	4.3219 (6.8222)	-2.0547 (1.8552)
Dummy- Wells in 2013	-10.3675 (9.7090)	0.6726 (2.6402)
Median Income	-0.0006* (0.0002)	-0.0001** (0.0000)
Income Squared	2.23e-09** (9.55e-10)	5.78e-10** (2.60e-10)
Total Population	0.0012 (0.0001)	0.0001** (0.0000)
Percent Rural Population	0.1129* (0.0366)	0.0249 (0.02845)
R-Squared	0.1093	0.0760
Number of Observations	167	167
Note: Standard errors are in parenthesis; Statistical significance is tested at the *1%, **5% and ***10% levels		

In the model above, I use dummy variables to account for the presence or lack of for oil and gas wells in the area in 2004 and 2013. None of the dummy variables are significant while median income shows the expected sign and relationship with oil and gas development.

Table X: Regression of Inter-temporal Difference using Dummies for Number of Wells as Independent Variables and "Cases by 1000 Population" as Dependent Variable		
Variable	Respiratory Illness	Fractures
Constant	-0.4135* (0.0633)	0.0125 (0.0307)
Dummy- Wells in both years	-0.0262 (0.0685)	0.0709 (0.0333)
Dummy- Wells in 2004	-0.2291 (0.2627)	-0.1904** (0.1275)
Dummy- Wells in 2013	-0.2389 (0.3731)	0.0102 (0.1810)
Median Income	-0.0000**	-7.85e-06**

	(6.36e-06)	(3.09e-06)
Income Squared	5.77e-11 (3.67e-11)	4.13e-11** (1.78e-11)
Percent Rural Population	0.0077** (0.0038)	0.0058* (0.0019)
R-Squared	0.0760	0.1254
Number of Observations	167	167
Note: Standard errors are in parenthesis; Statistical significance is tested at the *1%, **5% and ***10% levels		

This model is similar to the prior model, the difference is in the use of the number of cases per 1000 people in the population, accounting for population variability. Results show that oil and gas development is not correlated with the cases of respiratory illnesses in the area.

Table XI: Robustness Check for Respiratory Effects			
Variable	2004 Respiratory Illness	2013 Respiratory Illness	Inter-Temporal Difference (Respiratory Illness)
Constant	0.4506** (0.1937) (0.19089)**	0.1116 (0.1075) (0.0603)***	-0.1592* (0.0552) (0.0549)*
Wells/ Square-Mile	0.0008 (0.0040) (0.0007)	0.0129*** (0.0065) (0.0050)*	-0.0031 (0.0084) (0.0043)
Median Income	-0.0086 (0.0078) (0.0065)	0.0079* (0.0030) (0.0014)*	0.0048 (0.0040) (0.0024)***
Income Squared	0.0001 (0.0001) (0.0001)	-0.0001** (0.0000) (0.0000)*	-0.0000 (0.0000) (0.0000)***
Percent Rural Population	-0.0006 (0.0006) (0.0003)***	-0.0001 (0.0001) (0.0000)**	-0.0000 (0.0001) (8.30e-06)*
R-Squared	0.0083	0.0243	0.0036
Number of Observations	474	518	475
Note: Standard errors are in parenthesis; Statistical significance is tested at the *1%, **5% and 10% levels			

This model involves regressing all of the data elements without sorting for true zeros as was done with the other models. Using the data this way does not allow for comparison across because not all zip code areas in 2013 were present in 2004. Because 2004 is so much different form 2013 and coupled with the problem of too many missing variables in the data, these results cannot be reliable.

Table XII: Summary of Study Methodologies

Authors	Study Area	Data Source	Number of Obs.	Type of analysis	Total Cost Estimate	Cost per case Estimate
Schuler et al. (2006)	US-MN	Minnesota Department of Health and National Cost-per-case data	1,286,894-asthma specific observations	Top-down	$30.6 million	$1003
Davies and Hauge (2005)	US-WS	National Cost Estimates	Not Stated	Top-down	$406 million ($240 million-direct costs; $166.1 million-indirect costs)	Not applicable
Ladrigan et al. (2002)	United States	National Health Interview Survey(NHIS); CDC; Bureau of Labor Statistics; Health Care Financing Agency	Not Stated	Top-down	$2.0 billion*	Not applicable
Trasande and Liu (2011)	United States	Medical Expenditure Survey; Nationwide Inpatient Sample; Nationwide Emergency Department Survey; National Hospital Ambulatory Medical Care Survey/ National Ambulatory Medical Care Survey	Not Stated	Top-down	$2.2 billion*	Not applicable
Wang et al. (2005)	United States	Medical Expenditure Panel Survey; National Vital Statistics System	With asthma: 2,521,537	Regression -based	$1993.6	$791

			Without Asthma: 49,084,023			
Barnett and Nurmagambetov (2010)	United States	Medical Expenditure Panel Survey	206,851 (total pooled sample)	Regression -based	$3,259	(Direct Annual cost per case) $301, $93 (Annual Indirect costs per adult and per child)
Kamble and Bharmal (2009)	United States	Medical Expenditure Panel Survey	Not Stated	Regression -based	$37.2 billion	$1,004.6 (cost per adult) $2,0077.5 (cost per child)
Lozano et al. (1999)	United States	National Medical Expenditure Survey for 1987	With Asthma: 667 Without asthma: 6911	Bottom-up	Not applicable	$1129 (cost per child)
Smith et al. (1997)	United States	National Medical Expenditure Survey (NMEP) for 1987	Not Stated	Bottom-up	$5.8 billion	$2,584 (cost-per-patient) high-cost patients
Hasegawa et al. (2013)	United States	HCUP Kids Inpatient Database (KID)	Not Stated	Regression based	$1.27 billion-$1.59 billion	** $5940-$8410 (cost-per-discharge)
Weiss et al. (1992)	United States	National Centre for Health Statistics (NCHS); National Hospital Discharge Survey (NHDS); NAMCS; NHIS; NMCUES	Not Stated	Bottom-up	$6.2 billion	Not applicable

Note: Estimates shown above are specifically for asthma; in some cases authors report other illness estimates. Also, costs in this table have not been adjusted for inflation but are presented in its original author reported estimate. The however adjusted for inflation in the paper to help readers conceptualize the value of these costs in 2015.

* represents base estimates from a range of high-end and low-end estimates

** author looks at trends in asthma hospitalizations between 2000-2009; estimates represent increases in costs within that period