TRACKING THE LITTLE BLACK "RAIN" CLOUDS: AN ENVIRO-ECONOMIC ANALYSIS OF AMBIENT AIR POLLUTION EFFECTS ON PEDIATRIC ASTHMA

by

Jake Roberts Morgan

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Applied Economics

MONTANA STATE UNIVERSITY
Bozeman, Montana

January, 2012

APPROVAL

of a thesis submitted by

Jake Roberts Morgan

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to The Graduate School.

Dr. Anton Bekkerman

Approved for the Department of Department of Agricultural Economics and Economics

Dr. Wendy Stock

Approved for The Graduate School

Dr. Carl A. Fox

STATEMENT OF PERMISSION TO USE

ACKNOWLEDGEMENTS

Completing a thesis is no simple task, and I would like to first offer my gratitude to my committee chair, Dr. Anton Bekkerman, for guiding me through the process with patience and formal economic wisdom. The other committee members, Mark Anderson and Wendy Stock, answered my calls for additional committee meetings and provided insight from outside of the project which helped me clarify my message. Navigating the administrative labyrinths of actual thesis submission would have been impossible without the insider knowledge and advice from Jane Boyd and Jennifer Miller.

I am also thankful for the opportunity to collaborate with the Centers for Disease Control and Prevention on data and models used in this thesis. Jeanne Moorman, Paul Garbe, and the rest of the Air Pollution and Respiratory Health Branch were wonderful research hosts and helpful model critics. Thank you also to the Association of State and Territorial Health Officials for facilitating my presentation of results at the 2011 National Environmental Public Health Tracking Conference, and to the attendees there who offered practical advice for directing and applying my research.

Finally, the contribution of my parents and friends to this project may have not been in the form of economic or statistical advice, but their support proved to be just as important in completing this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Epidemiological studies routinely demonstrate a positive correlation between ambient air pollution and respiratory aggravation. Potential benefits from federal regulations to economic and physical health of individuals, however, are not well understood by either law makers or voters. As the EPA implements new regulations targeting SO_2 emissions, robust empirical analysis can frame the policy issue as one of statistical and economic analysis. The Asthma Call Back Survey extension of the Behavioral Risk Factor Surveillance System collects detailed data on symptoms, intervention methods, and demographics of individuals with asthma as well as associated ZIP-level spatial identification. The EPA's Acid Rain emissions database records detailed emission data for every power plant in the country. Together, these data can be used to link detailed asthma information with nearby plant emission levels, providing the foundation for an in-depth enviro-economic analysis of SO_2 effects on pediatric asthma aggravation and potential long term human capital investment. Using a zero-inflated negative binomial regression technique, the study estimates the effect that pollution has on days of missed school the provides insight into educational effects of emissions and presents the issue as one important to child development. The results and policy-oriented sensitivity analysis demonstrate the clear benefits of coal power plant pollution reduction, and suggest policy and education solutions which can mitigate negative asthma outcomes and promote educational attainments.

INTRODUCTION

During the 2009 State of the Union, President Barack Obama noted that the U.S. has one of the highest high school dropout rates among industrialized nations, proclaiming that "...dropping out of high school is no longer an option. It's not just quitting on yourself, it's quitting on your county" (Obama, 2009). Contemporary research supports this, finding that students' dropping out of high school are at elevated risk for "economic deprivation, detachment from school-based health service, and social, occupational, and marital problems in adulthood" (Kearney, 2008, 258). Kearney (2008) explain that youths with high absenteeism rates are at an elevated risk for permanent dropout. Moreover, absenteeism has also been linked with negative educational and human capital investment outcomes. This study examines how environmental factors, particularly ambient air pollution, can affect human capital investment through asthma aggravation.

It is difficult to directly measure human capital investment; however school absence has been used as a successful proxy in several other studies (as reviewed in Robert Wood Johnson Foundation, 2011). Analyzing school attendance is a common metric for linking chronic illness and educational outcomes, partly because survey questions of this information are available in several major health outcome datasets such as the National Health and Nutrition Examination Survey or Behavioral Risk Factor Surveillance System. Other studies (reviewed in the next chapter) also show that the number of missed school days is a credible variable to measure human capital development in children. That is, the primary form of human capital investment in children is schooling (specifically, the hours spent in the classroom). Missing a day of school implies a reduction in the ability of the child to enhance his or her human capital. Additionally, Moonie et al. (2006) suggest that absences can have a compounding effect, since missing one day can lead to lost information and lesser ability to be successful later in the semester, effectively decreasing the marginal returns to schooling.

The literature has also demonstrated successful empirical estimation using missed school days to proxy for human capital investment, especially in models incorporating chronic illness (Sexson and Madan-Swain, 1993).[1] One factor consistently found to increase absenteeism is chronic illness. Studies have shown that absence from school due to chronic illness interrupts the learning process of children compared to their healthy peers, depressing school performance and increasing dropout probability (Moonie et al., 2006). Asthma is the most prevalent chronic condition among children, affecting more than 6.3 million children in the United States alone (Robert Wood Johnson Foundation, 2011). Direct costs of managing the disease exceed $3.2 billion annually and indirect costs of lost parental productivity, familial stress, and government related costs are substantial (Wang, Zhong, and Wheeler, 2005; Levy, Winickoff, and Rigotti, 2011). The epidemic also affects the long term development of children, representing the leading cause in missed school days due to chronic illness (Robert Wood Johnson Foundation, 2011). Absenteeism, particularly due to medical reasons, decreases total human capital investment through lower adult earnings, decreased educational attainment, and increased behavioral disorder issues (Joe, Joe, and Rowley, 2009; Johnson et al., 2002).

Existing research has not fully identified all factors contributing to asthma in children. This study adds to the literature by developing an empirical bioeconomic analysis of pollution effects on missed school days due to asthma aggravation. This study uses unique data from the Centers for Disease Control and Prevention's Asthma

[1] Sexson and Madan-Swain (1993) explicitly cite four other studies which make use of the missing school days dependent variables.

Call Back Survey from 2007 through 2009, which includes an indicator of the number of days of school missed due to asthma aggravation. Various data-related and survey-related empirical issues are addressed by using a zero-inflated negative binomial model.

For several key reasons, this study is interested specifically in pediatric asthma incidence and aggravation. From an epidemiological perspective, children are more developmentally sensitive, so we would expect a more obvious effect of ambient air pollution on respiratory health. Understanding the effects on children in particular is also important for exploring long-run consequences to health compared to adults. It is rare for individuals to develop asthma beyond childhood, so an adult will either have asthma or not, while it may be possible to mitigate serious effects of asthma in children, either by reducing damage to developing lung tissue or instilling asthma control techniques. Combining the benefits analyzing asthma symptoms in children with the benefits of using days of school missed described by Moonie et al. (2006) and Sexson and Madan-Swain (1993) suggest that an empirical model focusing on missed school days in a model linking pollution output with human capital investment is prudent.

The underlying data generating process for missed school days is also important. The hypothesis is that increased air pollution results in asthma aggravation and missed school. If the days of school missed are being driven by something other than respiratory aggravation, our results would reflect that other driving force rather than our hypothesized effect. Ancillary regressions exploring the effect of pollution on asthma attacks confirm that pollution does affect absences through asthma aggravation and not some other cause.

This study's outcomes prove to be intuitive and have relevant policy implications. Point estimates reveal that increased pollution is associated both with increased number of asthma attacks and increased days of school missed due to asthma. In addition, a sensitivity analysis gives insight into the degree to which absenteeism can be reduced with pollution reduction. Specifically, the model predicts that with a 10% reduction in SO_2 levels would on average lead to a 2.2% decrease in school days missed, and a 30% reduction in pollution predicts a 6% decrease in missed school days. These results suggest several possible ways policy can be used to mitigate absenteeism. For example, policies targeted at continuing to decrease pollution can have substantial impacts on reducing absenteeism. Therefore, continued enforcement the Environmental Protection Agency's regulations targeting SO_2 emissions could have immediate and substantial effects on children's access to educational opportunities. The results also suggest that individual action could be effective in reducing absenteeism. For example, local educational programs for parents can increase awareness of the potential ambient air pollution risks, helping parents take steps to mitigate the severity of asthma response. Moonie et al. (2006) conclude that it is the severity of asthma responses, rather than just the presence of asthma symptoms that principally drive missed school days; therefore, even if parents could not limit all exposure to ambient air pollution, reducing the severity of exposure could yield dramatic results. In response to rising asthma-related medical costs, persistent high dropout rates, and heightened risk of absenteeism for asthmatic children, studies like this are important for investigating the relationships between asthma triggers and educational outcomes and suggesting effective policy action to combat the asthma problem and increase total social welfare.

REVIEW OF THE LITERATURE

Economic Literature

Asthma attacks can be triggered by ambient air pollution and the number of attacks over a given period can give an indication of relative magnitude of asthma severity. Missed days of school can be used to understand the way in which asthma episodes impact children's human capital development. Absenteeism has been shown to be correlated with decreased educational attainment, leading to reductions in adult socioeconomic and employment status (Joe, Joe and Rowley 2007). Among children who miss school because of medical reasons this effect can compounded (Joe, Joe, and Rowley, 2009). Missing a day of school can decrease the learning potential the following day while also making a child uncomfortable in a school environment. Medically, students with higher rates of absenteeism are more likely to exhibit anxiety and depressive disorders as well as losing interest or focus in class (Borrego et al., 2005). At the center of the concern over absenteeism is the demonstrated connection between missing school and dropping out (Kearney, 2008). The possibility of dropping out increases the potential negative effects of absenteeism to include violence, drug use, risky sexual practices, teen pregnancy, and future marital problems (Kearney, 2008).

Economic interest of asthma effects stems from both direct economic costs of the disease and its potential for impacting future human capital development. Wang, Zhong, and Wheeler (2005) detail many of these direct costs, calculating that medical expenditures topped $1 billion dollars for children in 1996 while also costing parents an average of $285 per child in lost productivity for parents due to missing work or transporting the child to a medical facility. Total costs of asthma have been estimated to exceed \$3.2 billion dollars a year (Weiss and Sullivan, 2001). Including payments for prescriptions, hospital visits, and doctor appoints this translates to an average of over \$401 per child per year (Wang, Zhong, and Wheeler, 2005). In addition to direct costs associated with managing asthma, there are numerous indirect costs that can impact long-term human capital development in children.

One such impact on human capital is on school absenteeism rates. In education, absenteeism has been shown to be a predictor of poor educational performance. Moreover, the severity of asthma closely tracks the number of days of school missed (Moonie et al., 2006). For example, Figure 2.1, reproduced from Moonie et al. (2006), tracks the rate of absence for non-asthmatics, asthmatics, and the rate of absence due to asthma absences for prospectively tracked asthma absences. The bottom two lines represent average missed school days among asthmatic and non-asthmatic schoolchildren. Based on our discussion of chronic illness, we expect that students suffering from asthma will overall have a higher rate of absence, which the figure confirms. Specific to our study, prospectively traced asthma absences based on absence due to asthma is significantly higher than other, non-disease related absence. This further motivates our hypothesis of the effects of asthma aggravation on absence. The figure demonstrates the large impact of asthma on missed days of school.

Asthma is a biologically complex condition, and consequently there have been studies focusing on many different factors which can trigger asthma aggravation. Xirasagar, Lin, and Liu (2006) and others have emphasized the importance of seasonality and pollen, while Aligne et al. (2000) that urban settings, where children are less exposed to pollen, are at elevated risk of aggravation. Breast feeding (Oddy, Peat, and de Klerk, 2002), obesity (Gilliland et al., 2003), and the presence of cats (Almqvist et al., 2001), among others, have also been demonstrated to affect pediatric asthma outcomes. These studies may be able to provide some guidance on

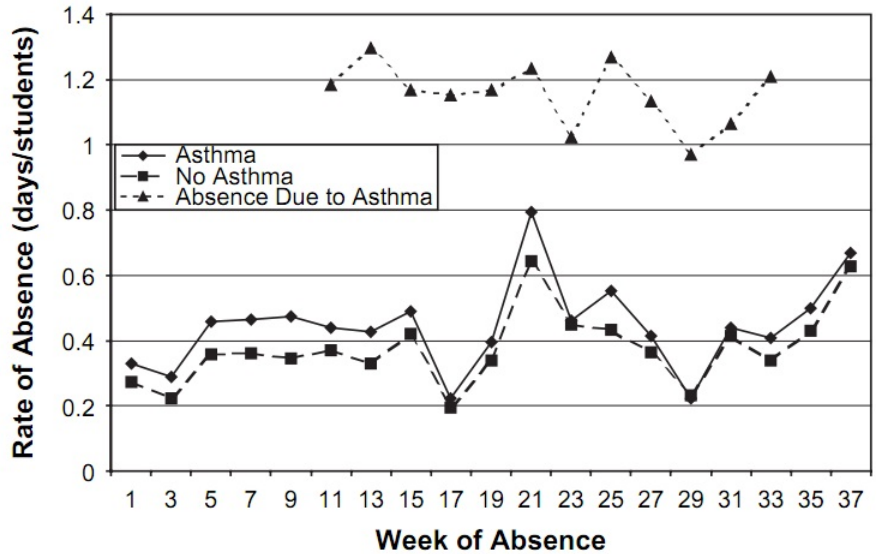

Figure 2.1: Absences Among Asthmatic Children

independent variable selection, but the connection to actionable policy change is less apparent compared to emissions.

Another factor that can affect asthma aggravation and adversely affect children's ability to attend school is ambient air pollution. Air pollution is a classic example of a negative externality, a condition when the costs of a particular activity are not fully internalized by the market participant. In the case of a polluter, this implies that the marginal cost of pollution that the polluter faces is less that the marginal social cost that pollution actually creates. Because property rights are not assigned to a common resource, as in the case of air pollution, the transactions costs are prohibitively high to allow for efficient allocation. While it may be possible to achieve pareto superior outcomes by assigning clear property rights to all resources (Coase, 1960), this is often politically infeasible. Furthermore, the number of economic actors in the clean air problem can complicate the issue, and suggests a more efficient, centralized, solution:

> when the gains or costs associated with particular interactions are not confined to a few parties, but instead are spread thinly over large numbers

> of individuals, then "high" transactions costs and "free rider" problems may be serious, even when utilizing the best of private property rights definitions, and some attenuation of private rights may be rationalized to achieve a more efficient solution to resource allocation problems When dealing with national defense and air pollution problems ... a rationalization for action by the state in the name of efficiency becomes available (Demsetz, 1979).

This study seeks to quantify the negative externality impacts of air pollution on children's educational development. By examining the relationship between pollution's impact on asthma aggravation and the resulting effects on school absenteeism we can better understand the long-term affects of pollution and better guide policy-maker action.

Epidemiological Literature

Understanding and appropriately modeling the impacts of pollution requires knowledge of epidemiological patterns and scientific mechanisms of asthma attacks. In the developmental phase, ages 0-18, the human body is immunologically less prepared to identify and cope with biological threats (Trasande, 2004). This is particularly true of the respiratory system. For example, acute increases in ambient air pollution have consistently been linked to asthma and other respiratory aggravations (Patel and Miller, 2009; O'Connor et al., 2008; Ho, 2010). Exposure to air pollution alters the normal process of lung development, air pollution can have lasting effects on respiratory health outcomes (Trasande, 2004). Therefore, studying childhood asthma is critical because exposure to asthma-inducing agents in the developmental phase can contribute to serious and long-term respiratory health problems.

Air quality studies often investigate the effects of emissions collected in two distinct methods: (1) emissions regulated by the Environmental Protection Agency (EPA);

and, (2) emission levels reported by power plant and monitor station records. Sulphur dioxide (SO_2) a pollutant that is recorded in both of these categories. The independent effect of SO_2 has become of particular interest, because the EPA mandates a limit on SO_2 under the Acid Rain Program (founded in 1990). Furthermore, on July 6, 2011, the EPA announced the Cross-State Air Pollution Rule (CSAR) targeting 27 states for substantial emissions reductions (Environmental Protection Agency, 2011). The rule specifically targets SO_2 from coal fire power plants and predicts substantial gains from emission reductions. The CSAR, which proposes a 73% reduction in SO_2 emissions, providing an impetus for analyzing quantifiable advantages to reducing pollution.

The effects of SO_2 on children's health have been investigated in several studies. Analyzing severity and frequency of wheezing in Italian hospitalized children, Orazzo et al. (2009) find a significant association between SO_2 and wheezing in the study cohort. These results are robust to multi-day lags, suggesting that SO_2 may have lingering health effects. Linares et al. (2010) find that SO_2 can also have additional, synergistic adverse effects when it interacts with other pollutants. For example, heavily polluted areas such as Mexico City, O_3 oxidizes SO_2 when it comes into contact with pulmonary tissue, acting as a respiratory irritant that triggers a dramatic airway response to allergens leading to negative respiratory symptoms. This indicates that in addition to having unique toxicological impacts, SO_2 can be especially harmful in combination with other pollutants (Linares et al., 2010).

Many studies tracking the effects of pollution on health outcomes published in epidemiological and economic scholarly publications use EPA monitor data, which tracks ambient air pollution wherever the monitor is located. These data are often preferred when the study's focus is on a narrow geographic region; such as ER visits in a single city (Zhu, Carlin, and Gelfand, 2003) or community clinic (Newcomb, 2006).

This allows researchers to gather pollution data on a small area not near a power plant and to analyze all sources of pollution at once, often yielding larger magnitude results than isolating a single contributor to overall pollution.

Relying on monitor data, however, both precludes a large national study (because the monitors are not uniformly dispersed throughout the nation) and does yield clear policy implications. Monitors do not distinguish where the ambient air pollution is coming from, so it is more difficult to target legislation or educational awareness at a specific harm. A distinct aspect of our study is that we employ source emission data.[1] Government agencies, such as the EPA, are particularly interested in source data because they are most conducive to crafting national policy. Power plants, for example, are tangible entities which can be regulated and measurably reduce pollution compared to ambient air pollution from an unknown source. Consequently, much of the literature using power plant emission data is published by or for the Environmental Protection Agency or similar organizations. For example, the Clean Air Task Force is one of the largest publishers and several of their studies suggest that power plant pollution, and SO_2 in particular, is closely related to negative health outcomes. Specifically, they conclude that a full 90% of the 25,000 estimated deaths due to pollution could be avoided by capping power plant pollution (Hill et al., 2001; Ledford and Clean Air Task Force, 2004). The results explored in these publications suggests that source pollution data can be useful in understanding stable health outcome estimates and providing a basis for policy analysis.

In these purely epidemiological studies, there are several factors that may limit the ability to draw broader economic inferences. For example, most studies use a narrowly defined unit of time (e.g., hourly or daily observations). Although this can provide

[1]In the literature source emissions refer to emissions from a source, such as a power plant or incinerator, distinct from monitor data.

detailed information about temporal pollution effects, there are several important drawbacks. First, studies using a shorter time interval (e.g., Andersen et al., 2008; Meng et al., 2010a; Neas et al., 1996) must also correct for a "plume effect," which describes the impacts of unsystematic changes in wind speeds and directions that can cause extreme data outliers in monitor data. One possible approach to help smooth these outlier effects and lessen the likelihood of outliers is using a more aggregated pollution measure. Also, while epidemiological studies generally include demographic and socioeconomic characteristics of a particular sample, often one or two of these variables are isolated as a variable of interest. By including and analyzing a full set of confounders, our study provides a more complete and intuitive understanding of the relationship between asthma, educational, and pollution outcomes.

ECONOMIC MODEL, EMPIRICAL SPECIFICATION, AND DATA DESCRIPTION

Theoretical Model

There exist several different measures of asthma outcomes in children, and choosing which of these variables to best describe the relationship between pollution and asthma outcomes is a matter of economic intuition and econometric theory. The economic question behind this investigation is first exploring the impact of those negative respiratory outcomes on human capital development and then investigating the mechanism by which pollution contributes the human capital effects. Intuitively, we will want variables to describe each of these parts, and those variables may not be the same. To judge the effect of asthma aggravation on human capital development, particularly education for children, a strong variable would be related to educational outcomes and show the same type of response to pollution as asthma attacks. An appropriate choice would be an educational variable more closely tracking pollution output and exhibiting more variation. Number of days of school missed due to asthma fits this profile and the literature on school attendance and human capital development invites our qualitative analysis. Kearney (2008) uses absenteeism to track dropout and academic achievement outcomes, and discuss the links of absenteeism to violence, substance use and decreased educational success. Moonie et al. (2006), Wang, Zhong, and Wheeler (2005), and others have investigated the effect that chronic health conditions, specifically asthma, have on educational outcomes and missed school days, finding unambiguously that asthma increases childhood risk for absenteeism. Researchers use missed days of school as an educational outcome variable in combination with health outcomes as a variable of interest both because of the wide availability of this indicator in health outcome surveys and because the variability of missed school tracks the severity of the health outcome. Our study links pollution to asthma aggravation and then to human capital investment, so an educational variable able to track changes in severity of asthma outcomes would be a good indicator.

Previous literature has documented the negative impact of pollution on respiratory outcomes, particularly asthma (Pope, Bates, and Raizenne, 1995; Neidell, 2004). In much of this literature, researchers focus on cost outcomes of asthma aggravation, but there is a gap in the literature directly relating coal power emissions to educational outcomes for children. To address this issue we would want data including a measure of missed school, pollution emissions data, and a variable to track the severity of asthma outcomes. If the analysis reveals that pollution has a significant effect in predicting both increased aggravation and missed school days, this would suggest that pollution is increasing missed school days through asthma aggravation and that a reduction of pollution exposure could curb both physiological asthma symptoms and student absenteeism.

Empirical Specification

We can use the data collected in the Behaviorial Risk Factor Surveillance System (BRFSS) Asthma Call Back Survey to empirically test the economic question presented in the theoretical model – to what extent and to what degree does ambient air pollution affect childhood educational outcomes. To measure absenteeism, the survey asks how many days of school the child missed in the last year due to asthma. This variable reports outcomes as whole numbers – 2 or 3 days of missed school, for example, rather than 2.5 days – so is a count variable. Similarly, the survey asks how many asthma attacks the child has experienced in the last three months and reports these values as counts. We can use these physiological outcomes to measure relative severity of asthma outcomes due to pollution. This variation is important when linking the outcomes to pollution, since our hypothesis is that more pollution will have a larger effect on asthma and missed school. As a robustness check, the data also include a binary indicator of whether or not a child had an attack in the last year which allows us to verify that the presence of coal power plant pollution increases the probability that a child suffers from an asthma attack.

The developed bioeconomic model considers pertinent epidemiological and economic factors affecting school absenteeism by children due to asthma aggravation. The model is represented by the function:

$$\begin{aligned} \textit{Asthma Outcome} &= f(\text{pollution, demographic variables, socioeconomic variables,} \\ &\quad \text{meteorological controls}) + \varepsilon_{it} \end{aligned} \tag{3.1}$$

where "pollution" represents a set of variables describing coal power plant emissions; "demographic variables" is a vector of racial, sex, and age characteristics of the adult respondent and relevant child; "socioeconomic variables" represent income and related indicators of class standing; "meteorological controls" here are primarily guided by the literature to include controls for weather patterns.

The error term, ε_{it},can be represented as

$$\varepsilon_{it} = \alpha_t + \xi_i + u_{it}$$

where α_t represents unobservable factors that vary across time for all individuals, ξ_i represents a time-invariant fixed effect, and u_{it} is a random error. Because pollution is not sedentary and confined to a well defined location, we follow existing literature to understand the behavior of spatial pollution dissemination and create a measure of ambient air pollution around a power plant.

"Pollution," as a single term, is difficult to incorporate in a detailed empirical study, as it is often representative of many complex and simultaneously occurring biologic and chemical processes. This is particularly the case when discussing ambient air pollution. There is, of course, no such thing as air pollution per se, and the commonly used measure of particulate matter (PM) is a combination sulfate, nitrate, and other organic and inorganic compounds which may differ in ratio at particular site of measurement (Reiss et al., 2007; Pope, Ezzati, and Dockery, 2009). Additionally, the strong correlation between particulate matter, its components sulfate and nitrate, and other toxins in the air presents a statistical problem when attempting to isolate the effects of a single contributing compound. This problem is compounded by the fact that these contributing compounds are often subject to biologic forces which cause them to change physical form, such as sulfur dioxide becoming ammonium sulfate when SO_2 oxidizes NO_2 in the atmosphere. Caution must be exercised, then, when conducting a study using only a select number of contributing parts of air pollution.[1]

Biological and chemical attributes of our pollutants of interest are also important to interpretation of the results. The co-movement of SO_2 and other pollutants, for example, gives us insight to what degree we can use SO_2 to proxy for the generalized effect of coal plant pollutants on human health. In using SO_2 as the pollutant of interest the presented model we are simply attempting to measure the generalized effect of air pollution from coal fired power plants rather than the effect of SO_2 in particular. The use of sulfur dioxide as a proxy, though, is not arbitrary. First, the

[1]Extending the example in Moolgavkar and Luebeck (1996), if two individuals ingest sugar laced with strychnine and sugar laced with potassium cyanide, respectively, it would be foolish to blame the sugar.

most comprehensive data set is the EPAs Acid Rain Program that collects emissions data on every plant in the county contains only three specific pollutant emissions: SO_2, NO_x, and Hg (Mercury). Of these, SO_2 is the most commonly used in general ambient air pollution studies as a pollutant of interest.[2] Next, specific to an interest in coal combustion, 67 percent of SO_2 production comes from (mostly) coal fuel combustion, compared to only 22 percent of NO_x, which carries policy implications for possible federal regulation of a specific emission (Reiss et al., 2007; U.S. Environmental Protection Agency, 2003). Third, the use of spatial analysis techniques has been found to affect different emissions in different ways. Taking particulate matter in general and SO_2 together in a multivariate regression, after implementing spatial analysis the PM coefficients were sharply attenuated, while SO_2 remained relatively unchanged by the correction (Krewski et al., 2000; Smith, 2003). Jerrett et al. (2005) conclude that analysis with PM is highly dependent on which subgroups are or are not included in the study, finding generally that controlling for certain indicators of socioeconomic status can greatly attenuate the resulting PM coefficient. Together these results imply that "the spatial analysis confirm that the association of PM with mortality is much less robust that is the association of SO_2 with mortality" (Reiss et al., 2007).

As the most robust to changes in model specifications and the strongest in association with mortality, sulfur dioxide provides the best pollution proxy measure. Other studies examining air pollution by tracking the gaseous forms of PM – namely SO_2 and NO_x – provide the foundation for the best method of analysis. Because air pollution is a complex concoction, and the oxidation of molecules within the pollution mix is also complex, it is important to rely on both epidemiologic facts and model intuition to dictate which pollutants are included in the regression model. Some

[2] Mercury has a large cohort of specific studies due in part to the strict regulation surrounding it and clear toxicity results, but is not actually used in general air pollution studies.

researchers (particularly Pope, Ezzati, and Dockery, 2009) suggest that including only a single pollutant of interest may be sufficient, particularly if levels of other confounding pollutants are low in the area surveyed. Because this study examines pollution effects on a national scope, the levels of pollutants vary across the country so this analysis clearly does not apply. Of course, it would be impossible, and perhaps redundant (based on the changing forms of molecules), to include every individual pollutant in the study. Because sulfur dioxide is a purely gaseous form (compared to PM, sulfate, and nitrate, for example), intuitively it makes sense to include other potential gaseous cofounders. In similar studies common cofactors include NO_x and, when available, O_3 (Novi, 2010). O_3, while it has been linked to a decrease in health respiratory functioning, oxidizes NO_x in the presence of hydrocarbons in the atmosphere, meaning, to some extent, the measures are mutually exclusive so that more O_3 will necessarily mean less NO_x *ceteris paribus.* So in the case where O_3 is not measured, as is the case in the primary data set for this thesis, the effect of O_3 will be reflected through the NO_x measure (Meng et al., 2010b). It is appropriate, then, to include NO_x in a model using SO_2 as a pollutant of interest.

Intuition and literature guide the choice of independent variables and expected signs of each. Among socioeconomic variables, we expect that more income would be associated with better health outcomes. Several authors have indicated that diet may play an important role in overall healthiness, but also asthma outcomes in particular (Rodriguez et al., 2002). The data do not provide direct controls for a childs diet, but higher income allows access to healthier foods, including fruits, vegetables, and organic products. In more general terms, income has been associated with better health outcomes even controlling for health insurance (as we do), so it makes sense that in our study higher income would also be related to better health outcomes. The presence of health insurance will result in better health outcomes, particularly because the negative asthma outcomes in this study, such as asthma attacks, can be controlled and minimized with medication as well as information provided by the doctor (Bousquet et al., 2005). It is expected that health insurance, which indicates a general "concern" about health and the accessibility of a doctor relative to those without insurance, also captures a part of an individuals attitude toward health. Combined with the direct effect of insurance decreasing serious problems associated with asthma, the effect of this variable should unambiguously be negatively signed, since individuals who obtain health insurance may be more cognizant of health in general. This is particularly true of parents obtaining health insurance for children. A young adult may not choose to insure themselves if they are healthy, but intuitively we expect a parent would want to protect their children.

Smoking is one of the most serious behaviorial aggravations of the respiratory system. This is well documented both among individuals who smoke and, more importantly for this study, among those subject to ancillary fumes from smoking – the so-called "second hand smoke" effect (Strachan and Cook, 1998). Beyond the biologic mechanisms that clearly indicate the harms of smoking, there exists a behavioral implication wrapped up in this variable as well. The knowledge about the dangers of smoking is so prevalent in modern society that continuing to smoke, especially in the presence of children, indicates a level of disregard of health. In our regression equation, we expect that a higher smoking value (i.e. if an individual smokes) would indicate both that the household may be exposed to the biologic harms of smoke and that attitudes in the household may be less healthy so that the sign of this variable would be negative. In fact, because the effects of smoking are so damaging, we expect the magnitude of this control to potentially be large relative to the other socioeconomic and demographic indicators.

Home environment has also been investigated as a potentially important demographic aspect of pediatric asthma outcomes (Tyra and Bryant-Stephens, 2009). Employment and marital status are both proxies we can use to broadly control for home environment of children. Employment measures general financial stability, and we expect that with greater stability individuals would be more likely to make long term investments in the future such as the health of their children. Similarly, marriage contributes toward familial stability, and given that divorce can be costly both in legally separating and losing a potential income stream this effect should be similar to employment.

In air pollution studies, the use or non-use of weather variables represents an issue of debate. Pope, Bates, and Raizenne (1995) call it a "critical issue," one which is a major point of discussion in time-series epidemiological models. Weather trends are often considered crucial to time-series analysis because of the period of time analyzed rather than the actual predictive contribution of the weather variables. In many air pollution studies where weather variables are taken into consideration, the time series analysis involves a limited geographic scope, such as a single school or a select number of cities, and pollution data on an hourly or daily level from local air monitors (For example Pope, Bates, and Raizenne, 1995; Andersen et al., 2008; McConnell et al., 2002). Because the time window of measurement is so narrow, we expect that time-series studies over a narrow geographic region are especially sensitive to significant outliers which, if large enough, would significantly skew the distribution and results. Indeed, because the source of pollution measurement is an independent monitor rather than the source point the results of these studies can pick up significant pollution outliers in the form of the "plume effect." That is, when a concentrated mass of pollution is moved across a potentially long distance by a wind current or storm system. This plume will then drop on a monitor, causing significant pollution readings. The elevated reading from this plume effect, however, does not reflect the true pollution in a particular geographic region, but rather weather patterns.

To address potential plume effects, studies use detailed weather controls including variables such as temperature, wind speed and direction, dew point, and pressure (Schwartz J, 1996; Levy, Greco, and Spengler, 2002). In addition to controlling for plume effects, these measures also control for variation in respiratory outcomes due to seasonality, such as increased lung aggravation due to increased allergens in the summer months. However, including too many weather variables has been shown to lead to significant overmodeling (Dockery and Schwartz, 1995). This thesis avoids overmodeling and the need to explicitly control for certain minor weather conditions in the following ways. First, this study uses pollution measures collected immediately after emission by a power plant. Compared to geographically dispersed pollution monitors, power plant readings record comparatively large numbers and are less affected by purely weather driven fluctuations. That is, all of the readings are directly linked to by-products of the plant's energy production and are trivially affected by minute weather changes. Using the pollution source as the primary unit of measurement minimizes the plume effect. Second, data are aggregated annually, smoothing out any potential outliers. The aggregation also implies that variation in pollution measures due to energy production differences on a particular day are less likely to skew estimation results.

Geo-spatial techniques used in this analysis further incorporate meteorological controls. To reflect the fact that air pollution spreads, the pollution indication used in the model for each ZIP code is a summation of power plant emissions from all other ZIP codes within a fifty mile radius. This geographic area is based on the findings from the Harvard School of Public Health indicating that the highest consistent deposits of sulfates occur within a fifty mile radius of a given coal fire power plant (Hill et al., 2001; Levy et al., 2000). Levy et al. (2000) implement a damage function model incorporating emissions data and detailed meteorological patterns to obtain this result, which reflects the area of highest concentration. While some EPA analyses have found that sulfate pollution can travel in excess of 750 miles (EPA 1995), this pattern of pollution spread would be difficult to model in the context of this thesis without advanced meteorological controls.

While including disaggregated or overly-narrow weather data is unnecessary, inclusion of broad weather trends can be beneficial. We consider the meteorological effects of temperature and precipitation which are two of the most important and most frequently controlled atmospheric processes in the literature (Pope, Bates, and Raizenne, 1995; Salmond and McKendry, 2009). Aggregating to the level of the pollution variables yields monthly averages for each variable based on the location of the measuring weather station.

Clearly, the potential drawback of limiting the radius to fifty miles, even with incorporating weather variables in the regression equation, is excluding individuals who live outside of this fifty mile radius but who may have been affected in some way by travelling plumes of power plant pollution. This exclusion implies that individuals who express asthma symptoms in the data set are reflected as living in an area not impacted by power plant pollution. Intuitively, assuming pollution does not become beneficial to respiratory health after a certain distance, this would produce estimates of the effect of power plant pollution which are a lower bound. If our results still support the hypothesis that more pollution is related to more days of school missed, then this implies that either the fifty mile radius is appropriate, or that the magnitude of the effect of pollution is large enough to overcome the dampening effect of the error.

Data Description

The data used in this investigation are collected from two primary sources: a restricted-access asthma callback survey administered by the Centers for Disease Control and Prevention as part of the Behavioral Risk Factor Surveillance System (BRFSS), and the Environmental Protection Agency's Acid Rain Program.

Public health tracking in the United States has a long history. For example, the Centers for Disease Control and Prevention was founded to combat malaria, and its current national and local efforts are focused on educating and protecting individuals from various diseases. In the 1980s, professional interest began to move from environmental causes of diseases to behaviorial effects and lifestyle choices on health outcomes (Kaprio and Koskenvuo, 1988a,b; Centers for Disease Control and Prevention, 2008). However, the federal government has limited programs aimed at investigating lifestyle risks and primarily relies on the National Center for Health Statistics for periodically tracking such information (Centers for Disease Control and Prevention, 2008).

To address this lack of information, feasibility surveys were conducted in a sample of states from 1981-83 culminating in the creation of the official Behavioral Risk Factor Surveillance System in 1984 (Centers for Disease Control and Prevention, 2008). The BRFSS is the largest telephone survey in the world and is a comprehensive effort on the part of the CDC to monitor "health risk behaviors, preventive health practices, and health care access primarily related to chronic disease and injury" throughout the United States and its territories (Centers for Disease Control and Prevention, 2008). The core questionnaire contains number of questions regarding pediatric asthma that the CDC uses to estimate asthma prevalence in each state. If a respondent indicates that (s)he or his/her children have asthma, then that respondent is invited to participate in an asthma follow-up survey administered two weeks after the core questionnaire. The pediatric asthma call-back survey contains a series of detailed questions regarding respiratory health and asthma symptoms as well as demographic information specifically relevant to asthma.

In addition to "core" questions developed by the CDC and asked in every survey, individual states have the option to develop and add their own questions when the survey is conducted in that state. The increasing focus on asthma in public health tracking was first observed at the state level, and added to the federal level survey in 1999. This addition corresponded to the formation of the CDC's National Asthma Control Program (NACP). The stated goal of the NACP is to "reduce the number of deaths, hospitalizations, emergency department visits, school days or workdays missed, and limitations on activity due to asthma" (Centers for Disease Control and Prevention, 2011). In partial fulfilment of this goal, an optional standardized asthma module was introduced for adults in 1999, and this module was incorporated into the BRFSS core questionnaire in 2000.

A broad question asking whether or not a child in the household had been diagnosed was incorporated as a module in 2002, but it was not until 2006 that a detailed separate asthma call back survey for children was implemented. The first full year of the study, 2006, had sparse participation relative to subsequent years and was omitted from this analysis. Therefore, this study uses data from that call-back data set from 2007 through 2009, the latest year available. All of these asthma outcome variables are count or binary, recording whole number of days of school missed, number of asthma attacks, or whether or not a child experienced an asthma attack. Table 3.1 shows the states participating in the call-back survey for each year of our analysis.

The EPA's Acid Rain Program tracks pollution output from every power plant in the United States, measuring emissions in yearly, monthly, daily, and hourly increments. The Acid Rain Program was founded in 1990 under the Clean Air Act, and requires power plants to report pollution output. The Acid Rain Program requires monitoring of emissions associated with coal production — sulfur dioxide (SO_2), nitrogen dioxide (NO_2), carbon dioxide (CO_2), and mercury(Hg). sulfur dioxide and nitrogen dioxide are emitted during the combustion of coal and prompt the oxidization to sulfuric acid (H_2SO_4), a major component of acid rain. Hg is also produced as a byproduct of coal production, although in substantially smaller quantities currently due to EPA regulations, and has been linked to neurological disorders in children (Zahir et al., 2005). The other major component of coal power plant emissions is CO_2, which while lacking in documented negative health effects compared to the other outputs is still of concern due to its controversial role in climate change. These data are a census of power plants and power plant emissions.

Lastly, meteorological variables are collected from the North American Regional Reanalysis (NARR) database, averaged across a particular month, and merged into the primary BRFSS dataset by ZIP code (Paciorek and Liu, 2009; Kim and Stockwell, 2008). These data are made available by a joint effort of the National Oceanic and Atmospheric Administration and Department of Commerce and uses a northern lamber conformal conic grid spatial model with extensive atmospheric measurements at 3-hr daily intervals (National Oceanic and Atmosheric Administration, 2011). The North American Regional Reanalysis draws on data from the National Oceanic and Atmospheric Administration's National Center for Environmental Prediction to deliver high resolution weather data (National Oceanic and Atmosheric Administration, 2011). Consistent with the literature, precipitation and temperature controls are included in our model.

Table 3.1: States Conducting the Child Asthma Call Back Survey

States	2007	2008	2009
Alaska	•		
Arizona	•	•	•
California	•	•	•
Connecticut	•	•	•
District of Columbia	•	•	•
Georgia	•	•	•
Hawaii	•	•	•
Illinois	•	•	•
Indiana	•	•	•
Iowa	•	•	•
Kansas	•	•	•
Louisiana			•
Maine	•	•	•
Maryland	•	•	•
Massachusetts	•	•	•
Michigan	•	•	•
Mississippi	•		
Missouri	•	•	
Montana	•	•	•
Nebraska	•	•	•
New Hampshire	•	•	
New Jersey		•	•
New Mexico	•	•	•
New York	•	•	•
North Dakota		•	•
Ohio	•	•	•
Oklahoma	•	•	•
Oregon	•	•	•
Pennsylvania	•		
Rhode Island		•	•
Texas	•	•	•
Utah	•	•	•
Vermont	•	•	•
Virginia		•	•
Washington			•
West Virginia	•	•	•
Wisconsin	•	•	•
Puerto Rico			•

Data-Related Modeling Issues

Table 3.2–Table 3.4 show descriptive statistics from the Asthma Call Back Survey, EPA power plant pollution measures, and meteorological variables between 2007 and 2009. The variables SO_2 and NO_x represent pollution output measures from the census of coal fire power plants in the United States. Questions relating to asthma outcomes and child demographics are reported by the adult survey respondent. Other demographic information, such as race, marital status, education, smoking, and insurance coverage are demographic characteristics of the adult respondent as reported by that respondent, and are included to control for the home environment of the child.

The summary statistics tables represent each of three regressions used to test the hypothesis that increased pollution leads to negative health and educational outcomes. In each table, the reported observations correspond to those used in each respective regression. The first noticeable feature is the similarity across the tables. The means of pollution are fairly constant, laying between 0.011 and 0.013 thousand tons of SO_2 and between 0.014 and 0.016 thousand tons of NO_x. Because most people in the United States do *not* live within fifty miles of a coal power plant, it is expected that these means are fairly low. Among power plants, the average SO_2 production is between 9,000 and 13,000 tons per year.[3] Socioeconomic and demographic variables are nearly identical, with similar proportions of the sample occupying a given class or racial group. Only seven of the independent variables – pollution parameters, weather parameters and interactions, and the child's age – are continuous, with most of the rest being binary responses. These variables are interpreted as simple percentages.

[3]This is based on SO_2 from Environmental Integrity Project (2006) and the number of coal plants nationwide – 315-380.

Table 3.2: Descriptive Statistics For Missed School Days

Variable	N	Mean	Std Dev	Min	Max
Days of School Missed	4,894	3.11	8.91	0	200
Pollution Variables					
SO_2 (in thousands of tons)	4,894	1.1E-2	0.23	0	11
NO_x (in thousands of tons)	4,894	1.4E-2	0.15	0	5
Socioeconomic Variables					
Income <$15k	4,894	0.04	0.19	0	1
Income $15-$35k	4,894	0.09	0.29	0	1
Income $35-$50k	4,894	0.07	0.25	0	1
Income >$50k	4,894	0.08	0.28	0	1
Employment Categories	4,894	2.31	2.17	1	9
Health Insurance	4,894	0.90	0.31	0	1
Demographic Variables					
Respondent Smokes Daily	4,894	0.14	0.35	0	1
White	4,894	0.83	0.37	0	1
Black	4,894	0.13	0.34	0	1
Asian	4,894	0.02	0.14	0	1
Native American	4,894	0.02	0.12	0	1
Other Race	4,894	0.07	0.26	0	1
Respondent Married	4,894	0.70	0.46	0	1
College Graduate	4,894	0.43	0.49	0	1
Child is a Boy	4,894	0.57	0.49	0	1
Child Age	4,894	3.75	5.16	0	18
Weather Variables and Interactions					
Precipitation (inches)	4,894	0.30	0.17	0	2
Temperature (degrees F)	4,894	54.06	16.48	-15	94
$SO_2 \times$ Temperature	4,894	0.55	12.59	0	727
$SO_2 \times$ Precipitation	4,894	0.00	0.08	0	4

Notes: All measures with minimum 0 and maximum 1 derive from binary survey questions. Employment is a categorical variable increasing with increasing levels of unemployment.

Table 3.3: Descriptive Statistics for Number of Asthma Attacks

Variable	N	Mean	Std Dev	Min	Max
Asthma Attack Quarterly Count	8,380	0.58	1.25	0	9
Pollution Variables					
SO_2 (in thousands of tons)	8,380	1.1E-02	0.21	0	11
NO_x (in thousands of tons)	8,380	1.5E-02	0.17	0	6
Socioeconomic Variables					
Income <$15k	8,380	0.03	0.18	0	1
Income $15-$35k	8,380	0.09	0.28	0	1
Income $35-$50k	8,380	0.06	0.24	0	1
Employment Categories	8,380	2.29	2.12	1	9
Health Insurance	8,380	0.90	0.30	0	1
Demographic Variables					
Respondent Smokes Daily	8,380	0.15	0.35	0	1
White	8,380	0.83	0.37	0	1
Black	8,380	0.13	0.33	0	1
Asian	8,380	0.02	0.15	0	1
Native American	8,380	0.02	0.13	0	1
Other Race	8,380	0.08	0.26	0	1
Respondent Married	8,380	0.71	0.45	0	1
College Graduate	8,380	0.43	0.49	0	1
Child is a Boy	8,380	0.60	0.49	0	1
Child Age	8,380	4.43	5.57	0	18
Weather Variables and Interactions					
Precipitation (inches)	8,380	0.30	0.17	0	2
Temperature (degrees F)	8,380	53.8	16.57	-15	94
$SO_2 \times$ Temperature	8,380	0.56	11.56	0	727
$SO_2 \times$ Precipitation	8,380	0.00	0.08	0	4

Notes: All measures with minimum 0 and maximum 1 derive from binary survey questions. Employment is a categorical variable increasing with increasing levels of unemployment.

Table 3.4: Descriptive Statistics for Binary Asthma Attack

Variable	N	Mean	Std Dev	Min	Max
Binary Asthma Outcome	5,346	0.60	0.49	0	1
Pollution Variables					
SO_2 (in thousands of tons)	5,346	1.3E-02	0.25	0	11
NO_x (in thousands of tons)	5,346	1.6E-02	0.19	0	6
Socioeconomic Variables					
Income <$15k	5,346	0.04	0.19	0	1
Income $15-$35k	5,346	0.09	0.29	0	1
Income $35-$50k	5,346	0.06	0.24	0	1
Employment Categories	5,346	2.37	2.18	1	9
Health Insurance	5,346	0.89	0.31	0	1
Demographic Variables					
Respondent Smokes Daily	5,346	0.14	0.35	0	1
White	8,380	0.83	0.36	0	1
Black	5,346	0.13	0.34	0	1
Asian	5,346	0.02	0.14	0	1
Native American	5,346	0.02	0.13	0	1
Other Race	5,346	0.07	0.26	0	1
Respondent Married	5,346	0.71	0.45	0	1
College Graduate	5,346	0.43	0.50	0	1
Child is a Boy	5,346	0.57	0.49	0	1
Child Age	5,346	4.24	5.43	0	18
Weather Variables and Interactions					
Precipitation (inches)	5,346	0.30	0.17	0	2
Temperature (degrees F)	5,346	53.7	16.55	-7	94
$SO_2 \times$ Temperature	5,346	0.69	14.04	0	727
$SO_2 \times$ Precipitation	5,346	0.00	0.09	0	4

Notes: All measures with minimum 0 and maximum 1 derive from binary survey questions. Employment is a categorical variable increasing with increasing levels of unemployment.

For instance, nearly 15% of the sample smokes daily and approximately 90% have health insurance.

The summary statistics also present several data concerns. The dependent variables of interest, days of missed school and number of asthma attacks, have standard deviations significantly above the mean and the range is wide. This implies potential overdispersion and excess zeros. Overdispersion in the data increases the deviance, which has the effect of overstating out t-values, generating misleading results. Too many zeros skew the data and can give misleading regression results. It may be the case that the child suffers zero missed days of school because of a factor we account for in the regression equation, so would be an appropriate zero value. However, it may also be the case that this asthmatic, for some unknown reason, is more resistant to manifestations of asthma which would cause him or her to miss a day of school. In this case our data would have "too many zeros" (Ridout, Demetrio, and Hinde, 1998).[4] Consequently, between overdispersion and too many zeros, we expect the data distribution to have a long tail, something that our histograms, Figures 3.1 and 3.2, confirm. Part of this is driven by possible outliers in the data, such as a student who misses 360 days of school or children who have had 100 asthma attacks in the last three months. Neither of these outcomes is particularly likely, especially because there is generally only 180-200 days in a school year and 90 days in a quarter. Even if some of the high results are possible, they are not representative of the true population and so skew the results. Appropriately, we truncate the highest 2% of data values.[5]

[4] Ridout differentiates these two types of zeros as structural versus sampling zeros.
[5] All results are qualitatively robust to using the full data set.

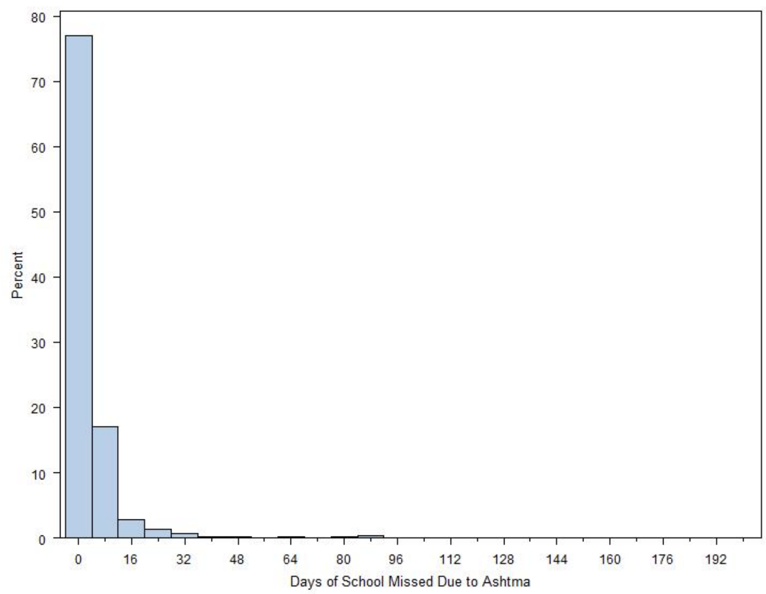

Figure 3.1: Missed School Days

Addressing Excess Zeros and Overdispersion

To proxy for educational attainment we use number of days of school missed due to asthma aggravation. This variable most closely relates to our investigation of human capital development. There are several attributes of these data which warrant special scrutiny. Chiefly, the reported responses are only whole days. So while it may be possible in the real world to pick up a child early due to asthma, meaning that missed school days would be continuous, it is not reported in this way in the data so is a count variable. Because the health outcome will not be continuous, instead clustered around whole numbers, the result will be inherently heteroskedastic compared to continuous data. The negative binomial distribution accommodates count data because it makes no assumption of continuous data instead using maximum likelihood procedures to estimate the parameters. This estimation technique is also distinct from other count data procedures, such as Poisson regressions, in that it has no assumptions regarding

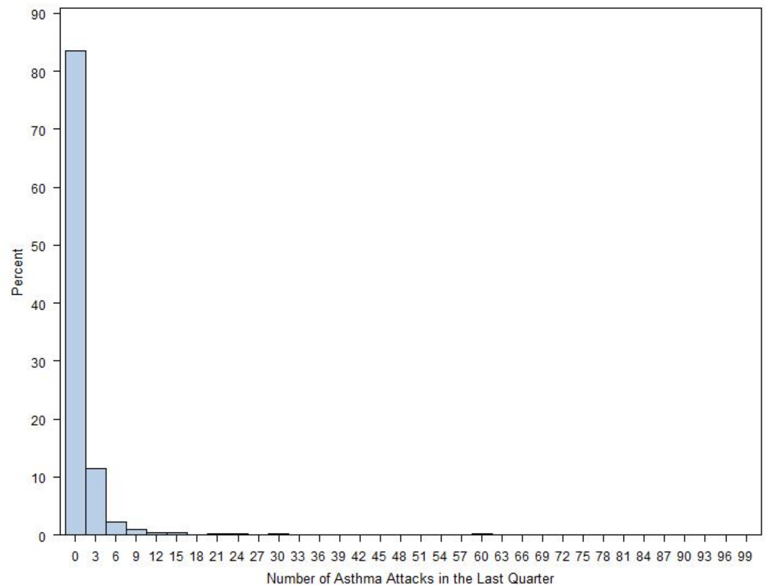

Figure 3.2: Number of Quarterly Asthma Attack

the deviance of the model, instead using the actual deviance of the model as the dispersion factor, automatically correcting for overdispersion.

A zero inflated count model corrects for this by diving responses into two categories – zero and non-zero. The so-called *zero-inflation probability*, ω is the probability that an individual would always have a zero, or "the probability of zero counts in excess of the frequency predicted by the underlying distribution" (SAS Institute, Inc., 2011). Zero inflating the negative binomial distribution gives the probability distribution,

$$Pr(Y = y) = \begin{cases} \omega + (1-\omega)(1+k\lambda) & for \quad y = 0 \\ (1-\omega)\frac{\Gamma(y+\frac{1}{k})}{\Gamma(y+1)\Gamma(\frac{1}{k})}\frac{(k\mu)^y}{(1+k\lambda)^{y+\frac{1}{k}}} & for \quad y = 1, 2, ... \end{cases}$$

which is used with probit link function (SAS Institute, Inc., 2011). In addition to the zero inflation probability, the distribution incorporates the underlying data distribution mean, λ, the gamma function Γ, and the negative binomial dispersion parameter

k, which corrects for overdispersion by estimating true dispersion instead of assuming a set parameter as the Poisson estimation strategy does. The associated mean, μ, and variance Y, are given by:

$$E(Y) = \mu = (1 - \omega)\lambda$$

$$Var(Y) = \mu + \left(\frac{\omega}{1 - \omega} + \frac{k}{1 - \omega} \right) \mu^2$$

At face value, the estimated coefficients yielded by the model are not particularly intuitive. With a zero inflated negative binomial regression model, we do not interpret the point estimates as marginal effects. Instead, we use the exponential of the estimate to calculate the factor by which the dependant variable changes for a one unit change in the independent variable. For out variable of interest, SO_2, we can find the factor by using the estimated coefficient. The values in the factor column in Table 4.1 use this exponential formulation to calculate the factor of each variable.

Identifying Selection Bias

When considering a geographically widespread region and analyzing only a subset of the population, systematic differences between included and non-included subjects in the study may result in sample selection bias. In the context of this thesis, the concern is that we are modeling the effect of SO_2 on asthma outcomes, but this effect is identified only by asthmatic children who live within fifty miles of a coal power plant. If households select into this population subset based on an unobservable characteristic in their utility function, then the sample would not be randomly distributed. In similar studies, fixed-effects methods are implemented to correct for this potential problem (as in Apelberg, Aoki, and Jaakkola, 2001). The presented estimating equation includes geographic and temporal fixed effects, but the form of the data allow for other methods targeting unobserved heterogeneity to be used as well.

Hernn, Hernndez-Daz, and Robins (2004) show that stratification of responses can provide an alternative to traditional fixed-effects methods. Structural stratification of data based on certain characteristics can control for unobserved heterogeneity similarly to fixed effects while offering the researcher more flexibility to design a stratification scheme suited to the data. Among epidemiological studies, selection bias is a common problem due to the prevalence of survey-based and voluntary data collection, which is closely correlated to socioeconomic status (SES) (Bornehag et al., 2006). Perhaps not surprisingly, SES is also closely tied to movement of individuals and selection of neighborhood residences, prompting researchers faced with this selection to extensively stratify based on economic status (Doff 2010). To correct for response rates, SES stratification is employed in the weighting scheme of the call-back BRFSS (Centers for Disease Control and Prevention, 2008). In addition, the complex survey weighting design also accounts for family size (number of children), which Doff (2010) also indicates can affect neighborhood selection. By using the CDC's complex weighting design, this analysis incorporates a stratification method directly applicable to neighborhood selection.

The theoretical reasoning behind the stratification strategy may be sound, but the geo-coded nature of the data allows for visual and statistical inspection of possible bias. Drastic differences in asthma outcomes in the two subsamples may indicate that individuals are indeed selecting into the different groups. Because we are not controlling for this self selection, our data are no longer random and estimated coefficient interpretations may be incorrect. In the context of this thesis, where the dependent variables of interest are health outcomes, a particular concern is that individuals at a higher risk of poor health outcomes, such as individuals with lower socioeconomic or educational attainment, will select into the treatment groups. For example, home prices near power plants may be lower relative to homes not near a power plant, inducing lower income individuals (who are also more likely for negative health outcomes) to be more likely to locate within a fifty mile radius.

To compare broadly those who live within a fifty mile radius and those who do not, we tabulated the weighted means of demographic variables including race, education, income, and smoking status. Doff (2010) and Bornehag et al. (2006) both suggest that these variables are some of the most important in determining neighborhood selection.[6]

Table 3.5: Comparing Demographic Means Between Power Plant and Non-Power Plant Residence

	Outside 50-mi. Radius	Within 50-mi. Radius
General demographics		
White	0.83	0.86
Black**	0.13	0.08
Asian	0.02	0.01
Native American	0.02	0.02
HS Grad	0.23	0.26
College Educated*	0.42	0.35
Income Brackets		
Less than $15k	0.04	0.05
$15-$35k	0.08	0.12
$35-$50k	0.06	0.08
More than $50k	0.09	0.11

* is significantly different at 10%, ** at 5%, *** at 1%

[6]While not as precise, plotting out response rates and power plant radius is a visual method for investigating selection bias. Ideally, such a plot would appear random, and not feature clusters of particular attributes around power plant locations. Indeed, graphing this on the ZIP level coincides with the statistical findings indicating there are no systematic differences between subgroups. However, because of the confidential nature of the data, that graphic is not shown.

To determine similarities between each group of observations we can use the mean for each variable of interest in each subpopulation to calculate a simple t-test for significant difference between the two groups. The t-test well take on the form

$$t = \frac{\bar{x}-u_0}{s/\sqrt{n}}$$

We can use the results of the t-test to determine whether or not the difference between the two means is significantly different from zero. If the difference is not significant, then the null hypothesis that the two subgroups are the same cannot be rejected. Because our treatment in whether or not an individual lives within fifty miles of a power plant, the test of significance will be between those who live outside that fifty mile radius, and those who live within it. Beginning with race characteristics from Table 3.5, the percentage of Caucasian respondents is 83% outside of the fifty mile radius and 86% within that radius. The rates for African Americans, Asians and Native Americans outside and inside the fifty mile radius are, respectively, 13% and 8%; 2% and 1%; and 2% and 2%. The results are similarly encouraging for the education variables. Outside of the fifty mile radius, 23% of the population has only a high school diploma and 42% have a college degree. Within the radius, 26% are high school graduates and 35% have a college degree.

Income is likely the demographic characteristic most correlated with residence choice. In the BRFSS data, income levels are separated into five distinct categories: below \$15,000, \$15,000 to \$35,000, \$35,000 to \$50,000, and above \$50,000. Using these groups it is possible to test if certain income brackets select to live either inside or outside the fifty mile radius. We are especially concerned about the extremes – very rich or very poor – since individuals in these categories are most likely to have their housing decisions influenced by income in the ways that would relate to selection bias and cause concern. For the population living within fifty miles of a power plant, 5% of the population has an income of less than $15,00 a year, 12% makes $15,000 to $35,000, 8% $35,000 to $50,000, and 11% of the population makes more than $50,000 a year. Among those not near a power plant, 4% make less than $15,000 per year, 8% make between $15,000 and $35,000, 6% make $35,000 – $50,000 and 9% make more than $50,000 per year. Using a simple t-test to compare these means yields the following p-values: 0.26, 0.16, 0.20, and 0.25.

These outcome indicate a general failure to reject the hypothesis that the difference between the means is zero. The only variable that shows marginal significant difference (more than 1 percent but less than 5) is African American race. This result, however, is slightly more unstable than some of the others because of the lower percentage of the black race in general. Also, the result is in the opposite direction we would expect, suggesting that more African Americans live outside the radius of a power plant. The chief concern would be that lower income (correlated with negative health outcomes) select to live near power plants, biasing the results. Moreover, the fact that every income bracket distinction is statistically the same particularly helps because selection based on income was one of the major concerns that prompted the selection analysis.

Selection Bias Robustness Check: Propensity Score Matching

Another method to investigate selection bias is Propensity Score Matching (PSM). Selection bias compromises the randomness of the data, implying that the estimates are no longer best and unbiased. PSM addresses this bias through a series of steps based on matching observations through analysis of estimated propensity scores.[7] First, a statistical software package estimates the conditional probability of an obser-

[7] The following steps are drawn from Guo and Fraser (2010).

vation receiving treatment. Using logistic regression, the process identifies covariates affecting selection bias and chooses a functional form to address them. After obtaining these conditional probabilities — propensities — the treated observations are matched with observations in a control group. The ultimate goal of this matching is to make treated and control groups as similar as possible. In this case, the treatment and control groups are matched on all of the relevant regression demographic and socioeconomic characteristics — income, race, sex, martial status, smoking, health coverage, and education. For example, the procedure will match a non-smoking married Asian family making over $50,000 a year who lives near a power plant to an identical household living outside of the fifty mile power plant radius. The regressions of interest are then performed on the matched sample and the output is compared to the full sample estimation. Results that are qualitatively the same suggest that there is evidence that full sample estimates are likely not affected by selection bias.

There are several different matching algorithms available to implement PSM. This study uses nearest neighbor matching, "the most straightforward matching estimator" (Caliendo and Kopeinig, 2008). Using this method, an observation from the control group is matched with a treated observation with the closest propensity score. Smith (2007) explains that if the primary purpose of implementing score matching is for classification rather than estimating structural coefficients, which is the fundamental goal of this procedure in this study. Moreover, a primary motivation for developing different matching techniques is non-binary treatment, which introduces further complexities into the model. Because our "treatment" for the purposes of selection bias is residence near a power plant — a binary question — a simple matching algorithm is sufficient (Caliendo and Kopeinig, 2008). For the purposes of the study, this straightforward method is an adequate model to check the robustness of our estimation strategy in the potential presence of selection bias. The model is also robust to alternative matching methods, such as Caliper and Kernal methods, which give similar results. Because there is no explicit reason to prefer these methods to nearest-neighbor, we will not discuss them further.

Because the propensity score matching technique is principally a robustness check, we are not looking for identical results as the full model, but general agreement on sign and magnitude of the variable of interest. To investigate this relationship we run two simple regressions of the effect of yearly pollution on days of school missed: the full sample and then the matched sample. The results are presented in the tables below. Not only are the signs consistent, but the magnitudes are similar as well.

Table 3.6: PSM Test

	Full Model Estimate	P-Value	Matched Model Estimate	P-Value
SO_2	0.67	0.048	0.41	0.091

The corresponding factor changes are 1.96 for the full model and 1.51 for the matched model, approximately a 20% difference. Fundamentally a robustness check, we are particularly interested in the significance of each of these results. These results only give further credence to our earlier finding of no statistically significant difference between the treatment and control group with respect to demographic variables.

Survey Modeling Issues

The BRFSS survey is a massive undertaking, in fact the largest telephone survey in the world, with 1.3 million household surveyed between 2007 and 2009. To more efficiently obtain such a large data set while still maintaining a level of randomness through random-dialing, the CDC exchanged simple random sampling in favor of disproportionate stratified random sampling (DSS).[8] In DSS, telephone numbers are divided into either "high density" or "medium density" strata depending on whether a number is publicly listed or not. To differentiate between high and medium density, the CDC developed "1+ blocks" – a computer-generated listing of 100 numerically consecutive telephone numbers containing at least one published household telephone number. All of the listed telephone numbers from the "1+ block" are the high density stratum and the number remaining in the "1+ block" after the listed numbers are removed are known as the medium density stratum (Centers for Disease Control and Prevention, 2005, 46-8). High density strata are sampled at a higher rate than medium density strata (1.5 times as often), allowing researchers to obtain a higher success rate, while still obtaining a representative sample. These density strata, however, are developed for calling efficiency and not the primary measure of geographic subpopulations (i.e., rural vs. urban). To capture this variation, each state is divided into regions and numbers in the state and numbers are randomly selected by region.

While DSS may be more efficient, it changes the probability that an individual is selected into the survey. Therefore, empirical results from using the unweighted data would not accurately reflect true population trends. In addition, DSS stratification can lead to disproportionate selection of a population relative to a state's population distribution, and nonresponse has been shown to be correlated with socioeconomic status and education (Schneider et al., 2010).[9] This could lead to an over- or under-identification of a particular subgroup. Because demographic factors such as income level are often important indicators of health outcomes (Pearlman et al., 2006), health surveys are especially subject to such bias.

---

[8]For more information on DSS and alternative sampling methods see Singh (2007).

[9]For example, if poorer individuals are systematically less represented in the survey because of decreased access to telephone service.

A final concern of national survey data is the possibility of unobserved heterogeneity. Intuitively, individuals in Georgia may be systematically different from individuals in Montana in ways in which are not captured in the regression. If this is the case, and if this difference affects asthma outcomes, then the model would suffer from omitted variable bias because we are not controlling for these differences. By explicitly including regional and time variables and through geographic stratification methods we can control for these unobserved differences.

Correction through Data Weighting

To correct for the non-random nature of the survey, the CDC developed several weights that can be used to develop a complex survey design analysis. In this study, three measures are used as survey weights. The first measure is the primary sampling unit (PSU) which is a cluster identifier for every surveyed individual belonging to a particular cluster. In the BRFSS, the PSU is defined by the Mitofsky-Waksberg sample design.[10] That is, the "1+ blocks" defined above consisting of 100 telephone numbers define the primary sampling unit and phone numbers are randomly selected from these blocks if the telephone number is from a household, the PSU is selected for further sampling; if not the PSU is rejected. Numbers from a particular PSU are dialed until three complete observations are obtained.[11] The second weight measure is a stratum variable, which identifies the strata (subsets of area code/prefix combination) to which a respondent belongs. The strata can be considered a geographic division between observations.

[10] See Casady (1993) for details of Mitofsky-Waksberg methods.

[11] In the context of the BRFSS, a complete observation means that an individual successfully completes the entire survey interview.

Lastly, there is a final weight applied to each respondent. This weight is constructed by the CDC and calculated as:

$$Child\ Weight = (strata\ weight) \times (1/(\#residential\ telephones)) \times$$
$$(\#\ children\ in\ household) \times (post - stratification\ weight) \qquad (3.2)$$

Child weight is the final weight assigned to each child to correct for the non-random sampling methods and is composed of several different measures. The strata weight controls for different probabilities of selection among different strata, and the post-stratification weight adjusts for non-coverage and non-response by dividing the number of children in age-by-gender or age-by-race-by-gender categories by the sum of the products of the preceding weights for the children in that same category. Residential telephones and the number of children in the household are included to control for non-response rates among those without telephones and to control for the probability that a child is selected to participate in the survey. These weights are applied to all empirical analysis in this study.

Addressing Unobserved Heterogeneity

A common characteristic of panel data sets is the problem of unobserved heterogeneity relating to geography. In this study, such heterogeneity is likely a result of unobserved geographic and temporal differences. If the heterogeneity is uncontrolled, results are likely affected by omitted variable bias (Wooldridge, 2006). Two popular methods used to address this issue are first-differencing of the data and a fixed-effects model, which uses dummy variables to control for unobserved effects (Biddle and Hamermesh, 1989; Card, 1999; Wooldridge, 2006). In a national study, unobserved geographic heterogeneity is of particular concern.

The combination of individual weights and strata variables address several potential sources of unobserved heterogeneity. Stratification helps investigate potential estimation bias. Because strata variables are unique for each different telephone area code region, including this variable explicitly controls for unobserved differences between geographic regions. To address issues of unobserved heterogeneity we include state fixed effects to control for regional differences and year fixed effects to address temporal issues. Additionally, stratification and weights are based on individual child age, education, and race difference, all of which further control for unobserved heterogeneity.

EMPIRICAL RESULTS

Under our hypothesis that increased pollution aggravates asthma and increases the number of days of school missed, we expect there to be a greater probability of being healthy if a participant lives farther away from a power plant. In our identification strategy, this would suggest that the control group – those who live outside of the fifty mile power plant radius – will have fewer missed days of school due to asthma compared to the treatment group. We can visually investigate our hypothesis by graphing the distribution of days of school missed, and checking for obvious differences. Figure 4.1 demonstrates the differences in the distribution of missed school days, showing that the probability of experiencing zero missed days is double in the control group. This suggests that those in the control group, away from a power plant, are healthier on average while more than 50% of children in the treatment group experience one or more days of absence due to asthma.

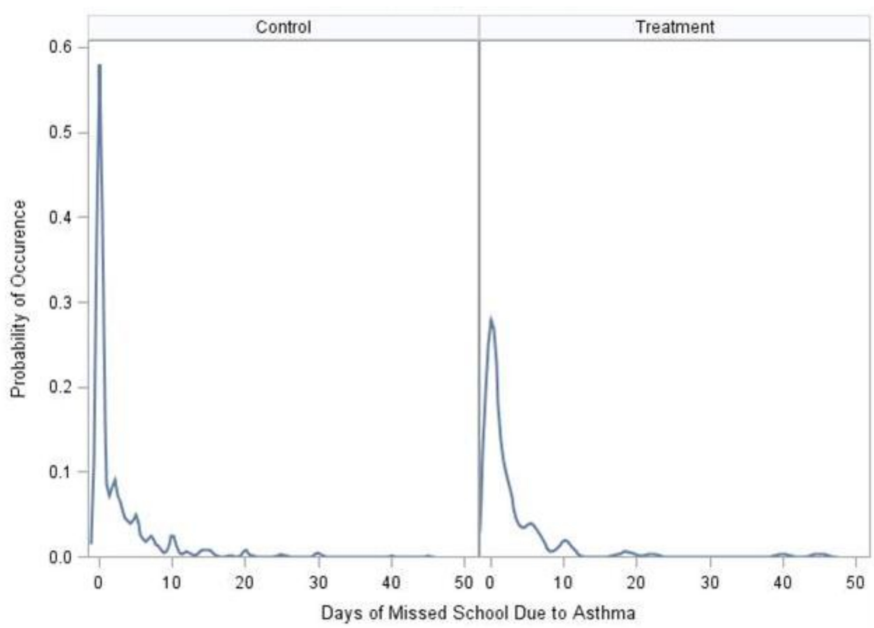

Figure 4.1: Distribution of Days of School Missed Due to Asthma

The empirical results agree with the initial hypothesis of the effect of ambient air pollution while in general agreeing in sign with the included demographic variables and reflecting the graphical representation of the treatment effect in Figure 4.1. Unambiguously, smoking, lower income, less health insurance and less employment all depress health outcomes in our model. Biologically, we know that black children are genetically more likely to suffer from asthma as are female children and both of these scientific facts are displayed in the results. The sign of our education variable is opposite in this specification, although not entirely unexpected. Indeed, this issue has confounded the field and is evident on a large scale with the increased asthma rate in first world (educated) countries compared to their uneducated counterparts in less developed nations. Some of this may be due to the "too clean" phenomenon studied by Braun-Fahrlander et al. (2002) and others. This explanation is based on hypersensitive parents, which is relevant to the question of missed school days. We expect the type of parenting (being over-protective) to be largely independent of the actually biologic mechanism behind asthma attacks, but of school attendance. If the "too-clean" hypothesis is correct, then it is reasonable to expect that some parents may be more prone to keep their children home from school because of respiratory symptoms than others, even if it is not a full-fledged asthma episode. The missed school days model was estimated using the zero inflated negative binomial specification and the results are presented in Table 4.1.

In this specification, increased levels of our pollutant of interest, SO_2 is significantly associated with missing more days of school. Using the exponential equation to convert negative binomial point estimates to interpretable factors,

$$SO_2 \quad \text{point estimate} = 0.67 \rightarrow \quad e^{0.67} \rightarrow \quad 1.96 \quad \text{factor},$$

Table 4.1: Analysis of Missed School Days

Parameter	Estimate	Factor	Standard Error	Pr > ChiSq
Pollution Variables				
SO_2 (in thousands)	0.67	1.96	0.34	0.05
NO_x (in thousands)	0.02	1.02	0.06	0.74
Socioeconomic Variables				
Income <$15k	0.27	1.30	0.04	<.01
Income $15-$35k	0.26	1.30	0.03	<.01
Income $35-$50k	-0.20	0.82	0.04	<.01
Employment Categories†	0.02	1.03	0.00	<.01
Health Insurance	-0.02	0.98	0.03	0.44
Demographic Variables				
Respondent Smokes Daily	0.06	1.06	0.02	0.01
Black	-0.07	0.93	0.02	0.00
Asian	-0.29	0.75	0.07	<.01
Native American	0.25	1.29	0.05	<.01
Other Race	0.24	1.27	0.03	<.01
Respondent Married	-0.02	0.90	0.02	0.30
College Graduate	0.04	1.04	0.02	0.03
Child is a Boy	-0.15	0.86	0.02	<.01
Child Age	-0.02	0.98	0.00	<.01
Weather Variables and Interactions				
Precipitation (inches)	-0.14	0.87	0.06	0.03
Temperature (degrees F)	0.00	1.00	0.00	0.05
$SO_2 \times$ Temperature	-0.02	0.98	0.01	0.04
$SO_2 \times$ Precipitation	-0.62	0.54	0.79	0.43

Notes: n=4894, state and year fixed effects included. McFadden pseudo-R2 = 0.20.
†*:* All model specifications are robust to inclusion of employment dummy variables. We use the combined categorical variable for notational and table brevity.

gives us a factor of 1.96.

Based on our discussion of meteorological variables in the theoretical section, if inclusion of such variables leads to major and unexpected changes in the estimated coefficients, or drastic loss of efficiency, we may be concerned about outlier bias or overparametization effects (Wooldridge, 2006). Large changes to the actual coefficient estimates would suggest that some weather outliers skewed the data, while loss of efficiency from increased variance is a symptom of including irrelevant variables. A robustness check without the two simple weather trend indicators that neither of these potential issues arises. The effect of including the weather variables is actually a *decrease* in variance across the regression variables, while the signs of the variables of interest remain unchanged. This suggests that the addition of weather trends is prudent and supports our more general results. The weather variables are significant when included in the regression, however, indicating that weather trends have important predictive power.

Trasande (2004) suggests that age can be an important factor in asthma development and outcomes, which is why we specifically control for it in the regression equation. If age is especially important, though, and different ages experience asthma in systematically different ways, then just including an indicator of age may be an inadequate specification. If the effect of age is non-linear, fit can be improved by including specifications of age which are also non-linear. The current specification is robust to the inclusion of a squared age term, which changes the point estimate of the variable of interest SO_2 only slightly from 0.67 to 0.79 and decreasing the p-value from 0.048 to 0.019, suggesting that the original model fits well with regard to age. We can also verify the robustness of the original specification by breaking up the sample by age and running separate regressions. Dividing the sample in half with regard to age – age 9 through 18 and age 8 and under – yields the same qualitative results. The coefficient on pollution is 2.28 (significant at the 1% level) for the eight-and-under group and 2.40 for the older sub-sample, although this is not significant. Neither of the sub-sampling results suggests that including all children in the same regression is incorrect, however. Also, Fitzpatrick, Grissmer, and Hastedt (2011) find that younger children in general are more susceptible to pollution, this result is intuitive. Moreover, they suggest that this is particularly relevant to education, as missed days of kindergarten can have a large effect on future educational outcomes, a finding which complements our hypothesis (Fitzpatrick, Grissmer, and Hastedt, 2011). If one of the models had revealed a reversal in sign of any variables of interest, it may have suggested that the age groups have such dichotomous models for missing school that running a combined model would have been inappropriate. Given that this is not the case, a single measure of age is retained for the final regression analysis.

As discussed in the theoretical chapter, the zero-inflated negative binomial models zero outcomes separately from the full model, so the analysis takes place in two separate stages. The first stage estimates the probability of a zero outcome, in our context the probability that a child experiences zero absences due to asthma, and then a negative binomial model predicts counts of days of school missed for children who are not certain zeros.[1] We are most interested in the change in number of days of school missed because this best tracks severity of asthma outcomes and changes in pollution, so provides the best basis for policy implications. For this reason, the results presented are the second stage, negative binomial results. However, the first stage, while not presented in full, provide some useful intuition. Most of the variables, including SO_2 are insignificant in the first stage, which is intuitive. There are many factors which influence asthma and asthma aggravation, and there is sufficient variation in

[1] In other words, the two stages separate out structural zeros in the first stage and include sampling zeros in the second stage.

number of missed school days over an entire year that the contribution of a single factor in determining the movement from no missed school to a positive number of missed days is difficult to isolate. Only some demographic variables, such as race, are significant, which is consistent with literature indicating that genetics play a large role in determining asthma severity. Among socioeconomic variables only extreme poverty, making less than fifteen thousand dollars a year, is significant, but signed opposite of what we would expect. This does not reveal any information regarding our empirical results per se, but does support our selection bias findings. Were selection bias playing a major role, we might expect that these socioeconomic indicators, instead of SO_2 would be contributing heavily to days of school missed, indicating that our second stage results were being driven by socioeconmics and not pollution. This is not the case, and the first stage results instead suggest the validity of our pollution analysis and the absence of powerful selection bias.

Because the variables in our study are in different units, comparing the magnitude of effect among the measures is not directly possible in the original form. Standardization of the variables allows us to compare relative importance of the variables in their effect on days of school missed due to asthma. To standardize a continuous variable, the mean is subtracted from the raw value and the resulting value is divided by the standard deviation in the following formulation:

$$V_{standardized} = \frac{V_{raw} - \bar{V}}{V_{SD}}, \quad \text{for each time period}$$

After running the primary regression using the standardized variables, the absolute value of the estimates gives the relative “importance” of the variable in affecting missed school days. Table 4.2 presents the results of the standardized regression of days of school missed due to asthma which are consistent with both our hypothesis and general intuition. SO_2 pollution ranks 9th among 18 independent variables, meaning it is not the most important factor affecting missed days of school due to asthma, yet still has a significant effect. From the literature on asthma formation, we do not expect pollution to be the most important variable affecting asthma aggravation, since the disease is heavily influenced by socioeconomic and demographic factors (Trasande, 2004). It makes sense, then, that income and race contribute the most, relatively, to days of school missed. Most telling are those variables which have a smaller relative effect compared to pollution, including health insurance, employment status, and parental educational attainment. In terms of educational opportunities curtailed due to asthma aggravation, then, this implies that public funds would be more effective targeted at reducing emissions rather than, say, increasing access to health insurance.

Sensitivity Analysis

Our regressions results are intuitive and give us economic insight into the relationship between pollution and asthma outcomes. To comment on practical policy prescriptions, though, it is useful to construct a counterfactual analysis. After implementing a methodologically legitimate model, we can run simulations to predict outcomes given certain changes to policy. This is particularly important because we do not expect that effects are necessarily constant across all levels of pollution. For example, we may expect that a 10% decrease of high pollution levels would have a larger impact than a 10% decrease from a smaller plant. To investigate this we use the predicted coefficients from our regression equation and then adjust SO_2 levels and analyze the change in the predicted values of the dependent variable. In Table 4.3 pollution levels are adjusted to 90%, 70%, 50%, and 10% of actual levels, and estimate predicted days of school missed based on the simulated pollution levels.

Table 4.2: Standardized Results of Days of School Missed

Standardized Variable	Mean	Std Dev	Chi-Square	Pr > ChiSq
Pollution Variables				
SO_2 (in thousands)	0.122	0.06	3.90	0.05
NO_x (in thousands)	0.003	0.01	0.11	0.74
Socioeconomic Variables				
Income <15k	0.266	0.04	49.8	<.01
Income 15-35k	0.262	0.03	87.2	<.01
Income 35-50k	-0.201	0.04	29.6	<.01
Employment Categories	0.025	0.00	44.6	<.01
Health Insurance	-0.020	0.03	0.58	0.44
Demographic Variables				
=1 if Respondent Smokes Daily	0.058	0.02	6.46	0.01
Black	-0.072	0.03	8.38	<.01
Asian	-0.288	0.07	15.1	<.01
Native American	0.252	0.06	21.1	<.01
Other Race	0.242	0.03	68.9	<.01
Respondent Married	-0.020	0.02	1.05	0.30
College Graduate	0.038	0.02	4.47	0.03
Child is a Boy	-0.147	0.02	81.5	<.01
Child Age	-0.086	0.02	23.6	<.01
Weather Variables and Interactions				
Precipitation	-0.023	0.11	4.69	0.03
Temperature	-0.019	0.01	3.97	0.05
$SO_2 \times$ Temperature	-0.042	0.08	4.15	0.041
$SO_2 \times$ Precipitation	-0.019	0.05	0.62	0.43

Notes: n=4894, state and year fixed effects included.

Table 4.3: Change in Predicted Missed School Days

Statistic	Avg. Days of School Missed	Avg. Days of School Missed, 10% Reduction	Avg. Days of School Missed, 30% Reduction	Avg. Days of School Missed, 50% Reduction	Avg. Days of School Missed, 90% Reduction
Minimum	0.024	0.010	0.002	0.0004	0
Maximum	11.725	11.724	11.723	11.722	11.7201
Mean	5.138	5.027	4.858	4.725	4.522
		%△ with 10% Reduction	%△ from 10% to 30% Reduction	%△ from 30% to 50% Reduction	%△ from 50% to 90% Reduction
Minimum		129	423	423	263
Maximum		0	0.01	0.01	0.02
Mean		2.21	3.48	2.81	4.49

Table 4.3 suggests several important things about our results. First, the intuition that the impact of changes in pollution levels may not be constant over all ranges of decrease is evident in these results. Reducing pollution levels by 10%, for example, decreases school absences by 2.2%, and decreasing by another 20% will reduce absences by an additional 1.3% for a total of 3.5% – diminishing marginal returns of pollution reduction. Next, despite the non-constant returns to pollution reduction, the direction of the change in the mean indicates that decreasing pollution is associated with less days of school missed due to asthma. At least, then, these simulations suggest that pollution has a clear negative impact on respiratory outcomes. At 90% reduction, average missed days decrease by approximately 15% from current levels. This suggests that reductions in pollution can yield positive returns to education, but also that asthma outcomes are influenced by many other factors as well.

This is particularly relevant in conjunction with implementation of the Cross-State Air Pollution Rule (CSAPR). The rule proposes that by 2014 annual SO_2 emissions be reduced by 6.4 million tons compared to 2005 levels – a 73% reduction, or about 8% a year – at a cost of $800 million annually.[2] The cost is largely reflective of plant upgrades, and under CSAPR 70% of coal fire power plants will be equipped with advanced pollution controls such as scrubbers. The benefit of additional school days is, however, more difficult to quantify. Hansen (2007) and Fitzpatrick, Grissmer, and Hastedt (2011) both demonstrate that, academically, being in school is an important predictor of standardized test performance. Specifically, five days of class attendance predicts an increase of test scores by 0.15 standard deviations (Hansen, 2007). This further contributes to the risk of dropout, which is perhaps one of the greatest eco-

[2]The reader is cautioned that this estimate has been attacked by both members of Congress and some in the coal energy industry. However, it is the most precise estimate available, and the estimate emerges from the same model the EPA uses to obtain the yearly SO_2 reduction estimate. For continuity, we use estimates from the same model.

nomic costs associated with missing school. On average, a high school drop out will cost the government more than $800,000 during his or her lifetime (Smink and Heilbrunn, 2005). Malone et al. (1997) point out that any estimation of direct costs of missing school will be "fraught" with burdensome assumptions, but the cost to caregivers who must stay home with a child on a missed school day is more clear. Wang, Zhong, and Wheeler (2005) calculate an expenditure of $268 per child with asthma for a total annual cost of $720 million dollars. The most extreme outcome of asthma aggravation is fatality due to asthma. In one year of the Wang, Zhong, and Wheeler (2005) study, a total of 211 school aged children died due to asthma, accounting for $265 million of lifetime earnings lost. While reducing pollution will not completely counter these costs, the sensitivity analysis indicates it will mitigate them. Indeed, the EPA estimates cost savings of the CSAPR of up to $280 *billion* annually, dwarfing the annual cost of $800 million.

Our motivation for investigating the effects of pollution on missed school days stems from the education economics literature, which suggests that missing more school days can have negative consequences for educational attainment and adult earning potential (Joe, Joe, and Rowley, 2009). In addition to educational attainment, students with higher rates of absenteeism are more likely to exhibit anxiety and depressive disorders (Borrego et al., 2005). Particularly relevant to education, children with higher rates of absenteeism are at greater risk for dropping out of school (Kearney, 2008). In one study, children with chronic conditions such as asthma a four percent less likely to be employed at age 33 compared to their healthy peers following statistically lower scores on standardized tests (Case, Fertig, and Paxson, 2005). Many of these negative outcomes appear in the economic education literature in the context of the consequences of premature school dropout, and excessive absence due to asthma is a strong predictor of dropping out of school (Moonie et al., 2006).

Moonie et al. (2006) go on to demonstrate that their results are strongly driven by asthma severity, so that the underlying severity of symptoms, rather than asthma alone, drives high dropout rates. Asthmatics on average already miss 2.48 more days of school compared to their non-asthmatic counterparts, strengthening our severity results (Wang, Zhong, and Wheeler, 2005). Because ambient air pollution acts as an asthma trigger, and our study closely tracks changes severity in aggravation, these finding are particularly applicable.

The sensitivity analysis can be used to support direct policy action promoting reductions in SO_2 levels by mandate or educational and tracking programs aimed at raising awareness of aggravation threats. Milwaukee, for example, established a system to track asthma-related hospital visits using an $870,000 grant (Robert Wood Johnson Foundation and Brown, 2006). The study identified a surprising number of individuals who were not diagnosed with asthma but who visited the emergency department for respiratory aggravation, thereby demonstrating a need for greater research in intermittent asthma (Robert Wood Johnson Foundation and Brown, 2006). Butz et al. (2011) use ambient air pollution data in conjunction with child respiratory outcomes to measure the benefits of in-home filtration systems. The study finds that obtaining an air filtering system can substantially reduce asthma symptoms among asthmatic children, indicating that there are relatively inexpensive methods for mitigating the effects of ambient air pollution. It is difficult to estimate all the benefits of educational or tracking programs because they are diverse among states and communities, but case studies indicate the potential benefits. Additionally, if the programs had no marginal benefit over the cost of the program, they would probably not be implemented in such large numbers (CDC, 2011).

Pollution abatement is expensive. Returning to Coase (1960) and the discussion of negative externalities, the problem with pollution is that it is a cost not clearly identified or assigned in the market system. Coase (1960) suggests assigning clear property rights, thereby bringing the externalities into the market and representing them more accurately in the price system. Placing a cap on the amount of SO_2 emitted is putting the cost primarily with the producers by way of scrubbers and other containment methods (although these may be passed to the consumer in some measure with higher utility costs), and is the strategy adopted by the EPA in the SO_2-limiting CSAPR. This strict cap assigns direct costs to producers, and is distinct from the previous method of SO_2 containment under the Acid Rain Program where a limited number of pollution credits were available for purchase to producers so that those companies with the highest willingness to pay for the ability to pollute were the ones to do so. However, the EPA found this inadequate in accounting for cost of the pollution to nearby states and did not contain overall pollution enough. Under the program pollution reduction occurred in the lowest cost areas, typically newer power plants where it is easier to update and add scrubbers, so sensitive ecological areas such as the Adirondacks and Great Smokies near older plants remain at risk (Office of Air and Radiation, 2002). These areas are environmentally important, but difficult to assign a value and are not a direct part of the market system because they are natural reserves rather than real estate property where the price could reflect the damage caused by pollution. Consequently, the costs were not fully accounted for in the acid rain cap and trade program. Pollution costs do not have to be borne in full by the pollution emitters, however. Because exposure to pollution can be minimized by individual behavior (Butz et al., 2011; Robert Wood Johnson Foundation and Brown, 2006) – for example playing indoors and using air filtration systems – another way to address the negative externalities would be to place the costs with the individuals exposed to, rather than emitting, pollutants.

Supporting Analysis: Physiological Evidence

The BRFSS survey offers two specific variables relating to asthma attack episodes. The first is a binary variable measuring whether or not the child has experienced an attack in the last year, giving a broad idea of the effect of pollution on aggravation. Because an individual's choices can often control and prevent asthma attacks, including medication use, diet, and environment, this binary indication is an important indication of the potential for elevated pollution levels to confound these preventative measures. The measure may also indicate that individuals (children and their parents) may not take pollution output in to consideration when engaging in preventative behaviors. We apply a survey logistic regression procedure both to properly account for the binary nature of the outcome and to follow the standard analysis techniques employed for preforming analysis on this particular dataset (Zahran, 2011).[3] We can represent the linear logistic model as:

$$logit(\pi) \equiv log(\tfrac{\pi}{1-\pi}) = \alpha + x\beta$$

The standard logistic regression assumes independent, identically distributed errors – an attribute clearly not applicable in survey sample data. With SAS we can specify a specific link function and include survey design variables such as the strata and weight variables discussed in the methodology section. Because of this, the procedure fits the model based on a common slopes cumulative model based on probabilities of response categories rather than individual probabilities (SAS Institute, Inc., 2011). The model results are detailed in Table 4.4.

[3]As mentioned previously, there are no published studies using the full child Asthma Call Back Survey in this particular way in scholarly journals. SAS analysis, however, is routinely applied internally at the Centers for Disease Control and Prevention. Since the agency designed and conducts the survey dataset, it is appropriate to look to their expertise in statistical analysis of the data.

Table 4.4: Analysis of Binary Asthma Attack Episode

Parameter	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Pollution Variables				
SO_2 (in thousands)	1.83	0.77	5.60	0.02
NO_x (in thousands)	-0.80	0.25	10.2	<.001
Socioeconomic Variables				
Income <$15k	0.18	0.39	0.20	0.66
Income $15-$35k	-0.22	0.22	1.05	0.31
Income $35-$50k	0.46	0.26	3.11	0.08
Employment Categories	0.03	0.03	0.77	0.38
Health Insurance	0.09	0.20	0.21	0.65
Demographic Variables				
Respondent Smokes Daily	0.23	0.19	1.49	0.22
Black	0.33	0.19	3.03	0.08
Asian	-0.65	0.44	2.17	0.14
Native American	-0.05	0.55	0.01	0.93
Other Race	-0.15	0.26	0.34	0.56
Respondent Married	0.03	0.15	0.05	0.82
College Graduate	-0.02	0.13	0.04	0.85
Child is a Boy	-0.15	0.12	1.44	0.23
Child Age	-0.004	0.00	4.66	0.03
Weather Variables and Interactions				
Precipitation (inches)	-0.33	0.52	0.40	0.53
Temperature (degrees F)	0.00	0.00	0.00	0.99
$SO_2 \times$ Temperature	-0.01	0.01	0.17	0.68
$SO_2 \times$ Precipitation	-3.09	2.16	2.05	0.15

n=5122. State and year fixed effects included.

The results from our first regression (Table 4.4) largely confirm our expectations about the effect of the demographic controls, but also the direction of the effect of pollution on asthma outcomes. The coefficients for our pollution measures are 1.83 for SO_2 and -0.80 for NO_x, both signs we expect. Recalling the discussion about the impact of omitting O_3, we expect nitrogen oxides to take on a negative sign. Sulfur dioxide takes on a positive sign, indicating that increasing the amount of coal power plant pollution in the air will increase the probability that a child suffers from at least one asthma attack. The logistic procedure models the log odds of a positive response – here if a child has experienced an asthma episode in the last year – as a linear combination of the independent variables. Applying this concept to our regression output, we can say that for a one unit change in SO_2 (an additional 1,000 tons of pollution) the difference in log-odds for a child having an asthma attack increase by 1.83. This explanation, however, is somewhat cumbersome. Following other health literature, we can also represent this as an odds ratio: the ratio of odds for a child having an asthma attack. Mathematically, we derive this through the following formulation:

$$g(X) \equiv log(\frac{Pr(asthmaattack=1|X)}{Pr(asthmaattack=0|X))})$$

We are particularly interested in the odds ratios for the pollution variables because of their interpretative power. Table presents these results.

Table 4.5: Odds Ratio Estimates for Binary Asthma Episode

Effect	Point Estimate
SO_2 (in thousands)	6.23
NO_x (in thousands)	0.48
Smoking	1.26

Our odds ratios reveal the same basic intuition as the logistic regression results: more SO_2 results in a higher probability of a child suffering from an asthma attack. Specifically, an asthmatic child living near a power plant is 6.23 times more likely to suffer from an asthma attack than an asthmatic child not living near a plant. This is made more striking by the fact that with proper control – environmental and medical – many asthmatics suffer from no attacks at all in a given year. In this data, a full 38% of respondents report that the child did not suffer from a single attack in the past 12 months, suggesting that ambient pollution from coal fire power plants may be enough to trigger an attack in otherwise healthy children.

Returning to the interpretation of the results in Table 4.4, our demographic controls conform to economic intuition. Because this binary outcome is so broad, and covers such a long period of time (1 year) it is more difficult to isolate individual contributors to the negative health outcome. In this regression the weather variables appear as insignificant, which makes sense in the context of a broad binary regression, where many of these weather effects "wash out" when averaging annually. The missed school regression tracks missed school days over the entire year as well, but the variability in this specification is much greater. Number of days missed approximates severity of aggravations, where the binary measure only tracks if the child has experienced at least one asthma attack. If one moderate day of weather provides the marginal aggravation to trigger an asthma attack, it will count as one for each dependent variable. If extreme weather, such as an exceptionally hot summer boosting pollen output, causes multiple attacks that summer, the binary measure will not pick that up, while the count will take this weather fluctuation into account by reflecting the increased severity of the respiratory aggravation. This intuition dictates that certain controls will be more important in some specifications than others. Still, though, those controls which are statistically significant give a complete picture of the relationship between asthma attacks and pollution. Using the odds ratio interpretation, children who live in a household with an adult who smokes are 1.3 times as likely to have suffered from at least one asthma attack in the last 12 months. Among other significant independent variables, older children were less likely to have experienced an attack compared to younger children, and increased precipitation can mitigate the adverse effects of pollution.

Measuring the number of asthma attacks is another check we can employ to support our findings regarding days of school missed. It is important that pollution affects missed school days *through* respiratory aggravation rather than some other ancillary cause. If children miss school because of an asthma check up at the doctors office, for example, our regression equation of interest might be measuring the responsibility of parents who live in a heavily polluted area rather than days of school missed because children were incapacitated by asthma. If this were the case, it would undercut our implications because the negative effects of missing school in this thesis and in the literature is premised on missing school due to medical illness, not preventative check ups. However, if pollution is demonstrated to trigger asthma attacks, we can better argue that the children who miss school are doing so for clear medical reasons. The relationship between pollution and asthma attacks is presented in Table 4.6.

Measuring the number of asthma attacks is similar to measuring the probability of having an asthma attack, so we expect similar, intuitive, results. However, because this dependent variable better tracks the severity of asthma aggravation we expect that our independent controls will more significantly predict asthma outcomes. Like the dependent variable of interest, days of missed school, we suspect that there may be many asthmatics who do not have an asthma attack over a particular quarter. Even if an attack occurs, many individuals can limit attacks with environmental control or medication use. Our histogram demonstrates that this is indeed the case, with 77%

Table 4.6: Analyzing Asthma Attack Counts

Parameter	Estimate	Factor	Standard Error	Pr > ChiSq
Pollution Variables				
SO_2 (in thousands)	1.39	4.02	0.82	0.089
NO_x (in thousands)	-0.13	0.87	0.14	0.339
Socioeconomic Variables				
Income <15k	-0.13	0.88	0.12	0.276
Income 15-35k	-0.06	0.95	0.08	0.466
Income 35-50k	0.16	1.18	0.07	0.026
Employment Categories	0.02	1.03	0.01	0.005
Health Insurance	0.03	1.03	0.07	0.604
Demographic Variables				
Respondent Smokes Daily	0.23	1.26	0.05	<.001
Black	-0.02	0.98	0.06	0.701
Asian	-0.18	0.83	0.19	0.336
Native American	-0.06	0.94	0.16	0.698
Other Race	-0.01	0.99	0.08	0.911
Respondent Married	-0.04	0.96	0.05	0.429
College Graduate	0.07	1.07	0.04	0.088
Child is a Boy	-0.01	0.99	0.04	0.766
Child Age	0.00	1.00	0.00	0.651
Weather Variables and Interactions				
Precipitation	0.00	1.00	0.15	0.977
Temperature	0.00	1.00	0.00	0.401
SO_2 × Temperature	-0.01	0.99	0.01	0.481
SO_2 × Precipitation	-4.13	0.02	2.47	0.094

n=7709. State and year fixed effects included.

of respondents reporting zero asthma attacks over the last three months. We expect this excess of zeros to over-disperse our data, biasing associated t-statistics upward. A zero-inflated negative binomial regression model, like the one used to model days of school missed, accounts for this excess of zeros, as well as the count-nature of the asthma attacks (attacks are recorded as whole numbers).

Our variable of interest, SO_2, is positively and significantly associated with asthma attacks, where 1,000 tons is associated with a four fold increase in additional asthma attacks over the last three months. Considering an average power plant with 10,000 tons of SO_2, our results imply that a 10% reduction in emissions would lead to four less asthma attacks over a given quarter.

Again, our demographic controls have the anticipated effect on asthma episodes. The presence of a smoking adult in the household is associated with a 1.23 factor increase in asthma attacks per quarter. Many of the remaining independent variables are insignificant, though increased unemployment is associated with more attacks, and, like other regression models, precipitation has a mitigating effect on pollution. While pollution is independently bad for respiratory health, rain can clear the air, so the more rain the better.

CONCLUSION

The Behavioral Risk Factor Surveillance System data can be used to approach many different economic questions, but the stated purpose of the dataset is partially rooted in creating public policy:

> States use BRFSS data to identify emerging health problems, establish and track health objectives, and develop and evaluate public health policies and programs. Many states also use BRFSS data to support health-related legislative efforts (Centers for Disease Control and Prevention, 2008).

This thesis ultimately supports this purpose, and we can use the empirical evidence to draw policy-related conclusions.

As with any economic analysis, it is important to be cautious when interpreting results. The first potential obstacle is the nature of the disease itself. The mechanisms of asthma, or even a universally accepted diagnostic definition, are controversial (Johnson et al., 2002). So while epidemiological literature plays an important role in independent variable selection, there is some degree of disagreement on the origin of asthma itself. However, our results study asthma *outcomes* rather than prevalence. Sources of aggravation, compared to sources of asthma, are widely studied, and there exists a far greater degree of consensus in the literature regarding causes of aggravation. This study follows the literature by focusing on aggravation, but we still must exercise caution in our analysis and broader implications for respiratory health. Empirically, the zero-inflated count models employed are consistent with the data, but it is unlikely we modelled reality perfectly (statistically our McFadden $R^2 = 0.20$). This is true for any economic model, but if something crucial was overlooked, our results might not be as precise or accurate as possible. Finally, although we have controlled for non-random survey data, we are intuitively concerned about possible sample self-selection in the data. Neither analysis of demographic means between treatment and control groups, nor propensity score matching analysis, revealed any evidence of major selection bias. We suspect, however, that health outcome analysis is very sensitive to income, race and location, so are still cautious about our results.

Despite these cautions, our results are intuitive, politically applicable, and they fulfill the mission of the BRFSS survey to analyze health outcomes and track changes therein. Our physiological analysis contributes to the literature tracking the relationship between asthma triggers such as ambient air pollution and asthma outcomes such as attacks. Asthma attacks and other physical symptoms of aggravation can become so severe that normal activities, such as school, are impeded. Indeed, our results suggest that increasing levels of SO_2 result in more days of missed school. Research has demonstrated that missing school, particularly because of medical reasons, can have substantial effects on educational attainment, socioeconomic status, and even serve as a proxy for increased morbidity (Levy, Winickoff, and Rigotti, 2011).

Table 5.1: Results Summarized

Variable	Factor	Odds Ratio
Days of School Missed Due to Asthma	1.96	–
Number of Asthma Attacks	4.02	–
Binary Asthma Attack	–	6.23

Although primarily used as support for our dependent variable of interest, missed school days, the results from the supporting physiological regressions have policy implications as well, as detailed in Table 5.1. Asthma represents a major healthcare burden, with the estimated cost of treating pediatric asthma over $3.2 billion annually, representing 70% premium in medical expenses for asthmatic versus non-asthmatic children (Robert Wood Johnson Foundation, 2011; Wang, Zhong, and Wheeler, 2005).

Fifty-one percent of these costs are either outpatient or hospital visits, both of which are strongly correlated with changes in asthma severity (Wang, Zhong, and Wheeler, 2005). This suggests that limiting potential aggravation and attacks can substantially affect costs associated with pediatric asthma. This is especially important in the context of missing school. Wang, Zhong, and Wheeler (2005) also estimate the costs of pediatric asthma due to adult guardian loss in productivity at $285 per child with asthma. In our sample the mean child age is about four years old (Table 3.2). Intuitively, if young children miss school because of a medical reason we may expect that a parent would often have to stay home with the ill child. This would imply that missed school days in conjunction with asthma episodes is particularly important with regard to loss of productivity on the part of the parent.

Both sets of regressions – the schooling regression of interest and the supporting physiological regressions – suggest that a reduction in coal power plant SO_2 pollution could have significant impacts in reducing asthma aggravation, improving educational outcomes, and cutting related medical costs. For the children, missing school is strongly correlated with early drop out, which has negative consequences for adult outcomes. More broadly, "school absenteeism may be used as a general marker of morbidity that is easily assessed using survey methods," so missing school due to asthma tracks the severity, and thus increasing costs, of the condition (Levy, Winickoff, and Rigotti, 2011). Indeed, while asthma affects a minority of children, the total costs for asthma are substantial, and many of these costs closely track severity of the condition (Wang, Zhong, and Wheeler, 2005). Policy action, whether in the form of direct legislative action, such as the Environmental Protection Agency's proposal to further reduce SO_2 in coal power plants, or through education of potential asthma trigger risks, may be able to mitigate a portion of these burdensome costs.

REFERENCES CITED

Aligne, A., P. Auinger, R. Byrd, and M. Weitzman. "Risk Factors for Pediatric Asthma: Contributions of Poverty, Race, and Urban Residence." *American Journal of Respiratory Critical Care Medicine* 162(2000):873–877.

Almqvist, C., M. Wickman, L. Perfetti, N. Berglind, A. Renstrm, M. Hedrn, K. Larsson, G. Hedlin, and P. Malmberg. "Worsening of Asthma in Children Allergic to Cats, after Indirect Exposure to Cat at School." *American Journal of Respiratory Critical Care Medicine* 163(2001):694–698.

Andersen, Z., P. Wahlin, O. Raaschou-Nielsen, M. Ketzel, T. Scheike, and S. Loft. "Size distribution and total number concentration of ultrafine and accumulation mode particles and hospital admissions in children and the elderly in Copenhagen, Denmark." *Occupation Environmental Medicine* 65(2008):458–66.

Apelberg, B. J., Y. Aoki, and J. J. Jaakkola. "Systematic review: Exposure to pets and risk of asthma and asthma-like symptoms." *Journal of Allergy and Clinical Immunology* 107(2001):455 – 460.

Biddle, J. E. and D. S. Hamermesh. "Sleep and the Allocation of Time." Nber working papers, National Bureau of Economic Research, Inc, 1989.

Bornehag, C.-G., J. Sundell, T. Sigsgaard, and S. Janson. "Potential self-selection bias in a nested case-control study on indoor environmental factors and their association with asthma and allergic symptoms among pre-school children." *Scandinavian Journal of Public Health* 34(2006):534–543.

Borrego, L., M. Csar, P. Leiria-Pinto, and J. Rosado-Pinto. "Prevalence of asthma in a Portuguese countryside town: repercussions on absenteeism and self-concept." *Allergologia et Immunopathologia* 33(2005):93 – 99.

Bousquet, J., S. Gaugris, V. S. Kocevar, Q. Zhang, D. D. Yin, P. G. Polos, and L. Bjermer. "Increased risk of asthma attacks and emergency visits among asthma patients with allergic rhinitis: a subgroup analysis of the improving asthma control trial." *Clinical I& Experimental Allergy* 35(2005):723–727.

Braun-Fahrlander, C., Allergy and Endotoxin Study Team, J. Riedler, U. Herz, W. Eder, M. Waser, L. Grize, S. Maisch, D. Carr, F. Gerlach, A. Bufe, R. Lauener, R. Schierl, H. Renz, D. Nowak, and von Mutius E. "Environmental exposure to endotoxin and its relation to asthma in school-age children." *The New England Journal of Medicine* 347(2002):869–77.

Butz, A. M., E. C. Matsui, P. Breysse, J. Curtin-Brosnan, P. Eggleston, G. Diette, D. Williams, J. Yuan, J. T. Bernert, and C. Rand. "A Randomized Trial of Air Cleaners and a Health Coach to Improve Indoor Air Quality for Inner-City Children With Asthma and Secondhand Smoke Exposure." *Archives of Pediatrics and Adolescent Medicine* 165(2011):741–748.

Caliendo, M. and S. Kopeinig. "Some Practical Guidance For The Implementation Of Propensity Score Matching." *Journal of Economic Surveys* 22(2008):31–72.

Card, D. "The causal effect of education on earnings." In O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics, Handbook of Labor Economics*, vol. 3, chap. 30. Elsevier, 1999, 1801–1863.

Casady, R. J. "Stratified Telephone Survey Designs." 1993.

Case, A., A. Fertig, and C. Paxson. "The lasting impact of childhood health and circumstance." *Journal of Health Economics* 24(2005):365 – 389.

Centers for Disease Control and Prevention. "Behavioral Risk Factor Surveillance System Users Guide." 2005. URL `http://www.cdc.gov/brfss/pdf/userguide.pdf`.

———. "About the BRFSS." 2008. URL `http://www.cdc.gov/brfss/about.htm`.

———. "CDC's National Asthma Control Program." 2011. URL `http://www.cdc.gov/asthma/nacp.htm`.

Coase, R. H. *The Problem of Social Cost*. Blackwell Publishing Ltd, 1960, 1–13.

Dockery, D. W. and J. Schwartz. "Particulate Air Pollution and Mortality: More Than the Philadelphia Story." *Epidemiology* 6(1995):pp. 629–632.

Doff, W. *Puzzling neighbourhood effects: Spatial selection, ethnic concentration and neighbourhood impacts*. IOS Press, 2010.

Environmental Integrity Project. "Dirty Kilowatts: Americas Most Polluting Power Plants." 2006. URL `http://www.dirtykilowatts.org/Dirty_Kilowatts.pdf`.

Environmental Protection Agency. "Cross-State Air Pollution Rule (CSAPR)." 2011. URL `http://www.epa.gov/airtransport/`.

Fitzpatrick, M. D., D. Grissmer, and S. Hastedt. "What a Difference a Day makes: Estimating Daily Learning Gains During Kindergarten and First Grade Using a Natural Experiment." *Economics of Education Review* 30(2011):269–279.

Gilliland, F. D., K. Berhane, T. Islam, R. McConnell, W. J. Gauderman, S. S. Gilliland, E. Avol, and J. M. Peters. "Obesity and the Risk of Newly Diagnosed Asthma in School-age Children." *American Journal of Epidemiology* 158(2003):406–415.

Guo, S. and M. Fraser. *Propensity Score Analysis: Statsitcal Methods and Applications*. SAGE Publications, 2010.

Hansen, B. "School Year Length and Student Performance: Quasi-Experimental Evidence." 2007.

Hernn, M., S. Hernndez-Daz, and J. Robins. "A structural approach to selection bias." *Epidemiology* 15(2004):615–25.

Hill, B. A., E. Baum, A. Hennen, I. Walton, and Clean Air Task Force. *Sulfur : sulfur emissions and midwest power plants*. Boston: Clean Air Task Force, 2001.

Ho, S.-M. "Environmental epigenetics of asthma: An update." *Journal of Allergy and Clinical Immunology* 126(2010):453–465.

Jerrett, M., R. Burnett, R. Ma, A. P. III, D. Krewski, B. Newbold, G. Thurston, Y. Shi, N. Finkelstein, E. Calle, and M. Thun. "Spatial Analysis of Air Pollution and Mortality in Los Angeles." *Epidemiology* 16(2005):727–36.

Joe, S., E. Joe, and L. L. Rowley. *Consequences of Physical Health and Mental Illness Risks for Academic Achievement in Grades K-12*. SAGE Publications, 2009.

Johnson, C. C., D. R. Ownby, E. M. Z. andSharon Hensley Alford, L. K. Williams, and C. L. M. Joseph. "Environmental Epidemiology of Pediatric Asthma and Allergy." *Epidemiologic Reviews* 24(2002):154 – 175.

Kaprio, J. and M. Koskenvuo. "A prospective study of psychological and socioeconomic characteristics, health behavior and morbidity in cigarette smokers prior to quitting compared to persistent smokers and non-smokers." *Journal of Clinical Epidemiology* 41(1988a):139 – 150.

———. "Relationship of education to major risk factors and death from coronary heart disease, cardiovascular diseases and all causes, Findings of three Chicago epidemiologic studies." *Liu, K and Cedres, LB and Stamler, J and Dyer, A and Stamler, R and Nanas, S and Berkson, DM and Paul, O and Lepper, M and Lindberg, HA and Marquardt, J and Stevens, E and Schoenberger, JA and Shekelle, RB and Collette, P and Shekelle, S and Garside, D* 41(1988b):139 – 150.

Kearney, C. "An Interdisciplinary Model of School Absenteeism in Youth to Inform Professional Practice and Public Policy." *Educational Psychology Review* 20(2008):257–282.

Kim, D. and W. R. Stockwell. "An online coupled meteorological and air quality modeling study of the effect of complex terrain on the regional transport and transformation of air pollutants over the Western United States." *Atmospheric Environment* 42(2008):4006–4021.

Krewski, D., R. T. Burnett, M. S. Goldberg, K. Hoover, J. Siemiatycki, M. Jerrett, M. Abrahamowicz, and W. H. White. "Reanalysis of the Harvard Six Cities study and the American Cancer Society study of particulate air pollution. Part II. sensitivity analysis." Tech. rep., 2000.

Ledford, A. and Clean Air Task Force. *Dirty Air, Dirty Power: Mortality and Health Damage Due to Air Pollution from Power Plants.* Boston: Clean Air Task Force, 2004.

Levy, D. E., J. P. Winickoff, and N. A. Rigotti. "School Absenteeism Among Children Living With Smokers." *Pediatrics* 128(2011).

Levy, J., S. Greco, and J. D. Spengler. "The Importance of Population Susceptibility for Air Pollution Risk Assessment: A Case Study of Power Plants near Washington, DC." *Environmental Health Perspectives* 110(2002):1253–1260.

Levy, J., J. Spengler, D. Hlinka, and D. Sullivan. "Estimated Public Health Impacts of Criteria Pollutant Air Emissions from Nine Fossil-Fueled Power Plants in Illinois." Harvard school of public health paper, 2000.

Linares, B., J. Guizar, N. Amador, A. Garcia, V. Miranda, J. Perez, and R. Chapela. "Impact of air pollution on pulmonary function and respiratory symptoms in children. Longitudinal repeated-measures study." *BMC Pulmonary Medicine* 10(2010):1–9.

Malone, D. C., K. A. Lawson, D. H. Smith, H. M. Arrighi, and C. Battista. "A cost of illness study of allergic rhinitis inthe United States." *Journal of Allergy and Clinical Immunology* 99(1997):22 – 27.

McConnell, R., K. Berhane, F. Gilliland, S. J. London, T. Islam, W. J. Gauderman, E. Avol, H. G. Margolis, and J. M. Peters. "Asthma in exercising children exposed to ozone: a cohort study." *The Lancet* 359(2002):386–91.

Meng, Y., R. Rull, M. Wilhelm, C. Lombardi, J. Balmes, and B. Ritz. "Outdoor air pollution and uncontrolled asthma in the San Joaquin Valley, California." *Journal of Epidemiological Community Health* 64(2010a):142–7.

Meng, Y., R. P. Rull, M. Wilhelm, C. Lombardi, J. Balmes, and B. Ritz. "Outdoor air pollution and uncontrolled asthma in the San Joaquin Valley, California." *Journal of Epidemiology and Community Health* 64(2010b):142–147.

Moolgavkar, S. H. and E. G. Luebeck. "A Critical Review of the Evidence on Particulate Air Pollution and Mortality." *Epidemiology* 7(1996):420–428.

Moonie, S. A., D. A. Sterlin, L. Figgs, and M. Castro. "Asthma Status and Severity Affects Missed School Days." *Journal of School Health* 76(2006):18 – 24.

National Oceanic and Atmosheric Administration. "NCEP North American Regional Reanalysis: NARR." 2011. URL `http://www.esrl.noaa.gov/psd/data/gridded/data.narr.html`.

Neas, L. M., D. W. Dockery, H. Burge, p. Koutrakis, and F. E. Speizer. "Fungus Spores, Air Pollutants, and Other Determinants of Peak Expiratory Flow Rate in Children." *American Journal of Epidemiology* 143(1996):797–807.

Neidell, M. J. "Air pollution, health, and socio-economic status: the effect of outdoor air quality on childhood asthma." *Journal of Health Economics* 23(2004):1209–1236.

Newcomb, P. "Results of an Asthma Disease Management Program in an Urban Pediatric Community Clinic." *Journal for Specialists in Pediatric Nursing* 11(2006):178–188.

Novi, C. D. "The influence of traffic-related pollution on individuals' life-style: results from the BRFSS." *Health Economics* 19(2010):1318–44.

Obama, B. "State of the Union Address." Presented as the 219th Annual State of the Union Address, Washington, DC, 2009.

O'Connor, G. T., L. Neas, B. Vaughn, M. Kattan, H. Mitchell, E. F. Crain, R. E. III, R. Gruchalla, W. Morgan, J. Stout, G. K. Adams, and M. Lippmann. "Acute respiratory health effects of air pollution on children with asthma in US inner cities." *Journal of Allergy and Clinical Immunology* 121(2008):1133–1139.e1.

Oddy, W. H., J. K. Peat, and N. H. de Klerk. "Maternal asthma, infant feeding, and the risk of asthma in childhood." *Journal of Allergy and Clinical Immunology* 110(2002):65 – 67.

Office of Air and Radiation. "Clearing the Air: The Facts About Capping and Trading Emissions." Epa-430f-02-009, 2002.

Orazzo, F., L. Nespoli, K. Ito, D. Tassinari, D. Giardina, M. Funis, A. Cecchi, C. Trapani, G. Forgeschi, M. Vignini, L. Nosetti, S. Pigna, and A. Zanobetti. "Air Pollution, Aeroallergens, and Emergency Room Visits for Acute Respiratory Diseases and Gastroenteric Disorders among Young Children in Six Italian Cities." *Air Pollution, Aeroallergens, and Emergency Room Visits for Acute Respiratory Diseases and Gastroenteric Disorders among Young Children in Six Italian Cities* 117(2009):1780–85.

Paciorek, C. J. and Y. Liu. "Limitations of Remotely Sensed Aerosol as a Spatial Proxy for Fine Particulate Matter." *Environmental Health Perspectives* 117(2009):904–909.

Patel, M. and R. Miller. “Air pollution and childhood asthma: recent advances and future directions.” *Current Opinions in Pediatrics* 21(2009):235–42.

Pearlman, D. N., S. Zierler, S. Meersman, H. K. Kim, S. I. Viner-Brown, and C. Caron. “Race disparities in childhood asthma: does where you live matter?” *Journal of the National Medical Association* (2006).

Pope, C. A., M. Ezzati, and D. W. Dockery. “Fine-Particulate Air Pollution and Life Expectancy in the United States.” *New England Journal of Medicine* 360(2009):376–386.

Pope, I., D. V. Bates, and M. E. Raizenne. “Health Effects of Particulate Air Pollution: Time for Reassessment?” *Environmental Health Perspectives* 103(1995):472–480.

Reiss, R., E. L. Anderson, C. E. Cross, G. Hidy, D. Hoel, R. McClellan, and S. Moolgavkar. “Evidence of Health Impacts of Sulfate-and Nitrate-Containing Particles in Ambient Air.” *Inhalation Toxicology* 19(2007):419–449.

Ridout, M., C. Demetrio, and J. Hinde. “Models for Count Data with Many Zeros.” 1998.

Robert Wood Johnson Foundation. “Models for Advancing Asthma Care: Asthma Burden.” 2011. URL `http://www.pediatricasthma.org/about/asthma_burden`.

Robert Wood Johnson Foundation and M. H. Brown. “Milwaukee Hospital Develops Web-Based Pediatric Asthma Tracking System and Educational Program for Emergency Department Visits.” Grant results, 2006.

Rodriguez, M. A., M. A. Winkleby, D. Ahn, J. Sundquist, and H. C. Kraemer. “Identification of Population Subgroups of Children and Adolescents With High Asthma Prevalence: Findings From the Third National Health and Nutrition Examination Survey.” *Archives Pediatric Adolescent Medicine* 156(2002):269–275.

Salmond, J. and I. McKendry. *Air Quality in Urban Environments*, chap. Influences of Meteorology on Air Pollution Concentrations and Processes in Urban Areas. 2009.

SAS Institute, Inc. *SAS/STAT 9.3 Users Guide*. Cary, NC: SAS Institute Inc, 2011.

Schneider, K. L., M. A. Clark, W. Rakowski, and K. L. Lapane. “Evaluating the impact of non-response bias in the Behavioral Risk Factor Surveillance System (BRFSS).” *Journal of Epidemiology and Community Health* (2010).

Schwartz J, N. L. “Is daily mortality associated specifically with fine particles?” *Journal of Air Waste Management Association* 46(1996):927–39.

Sexson, S. B. and A. Madan-Swain. "School Reentry for the Child with Chronic Illness." *Journal of Learning Disabilities* 26(1993):115–137.

Singh, K. *Quantitative Social Research Methods*. SAGE, 2007.

Smink, J. and J. Heilbrunn. "Legal and Economic Implications of Truancy." 2005.

Smith, J. P. "The Impact of Socioeconomic Status on Health over the Life-Course." *Journal of Human Resources* 42(2007):739–64.

Smith, R. L. "Invited Commentary: Timescale-dependent Mortality Effects of Air Pollution." *American Journal of Epidemiology* 157(2003):1066–70.

Strachan, D. P. and D. G. Cook. "Parental smoking and childhood asthma: longitudinal and case-control studies." *Thorax* 53(1998):204–212.

Trasande, L. "The role of air pollution in asthma and other pediatric morbidities." *Journal of Allergy and Clinical Immunology* 115(2004):689–699.

Tyra and Bryant-Stephens. "Asthma disparities in urban environments." *Journal of Allergy and Clinical Immunology* 123(2009):1199 – 1206.

U.S. Environmental Protection Agency. "National air quality and emissions trends report, 2003 special studies edition." Epa/454/r-03-005, 2003.

Wang, L. Y., Y. Zhong, and L. Wheeler. "Direct and Indirect Costs of Asthma in School-Age Children." *Preventatitive Chronic Diseases* 2(2005).

Weiss, K. B. and S. D. Sullivan. "The health economics of asthma and rhinitis. I. Assessing the economic impact." *Journal of Allergy and Clinical Immunology* 107(2001):3 – 8.

Wooldridge, J. *Introductory Econometrics : A Modern Approach*. Thomson South-Western, 2006.

Xirasagar, S., H.-C. Lin, and T.-C. Liu. "Seasonality in pediatric asthma admissions: the role of climate and environmental factors." *European Journal of Pediatrics* 165(2006).

Zahir, F., S. J. Rizwi, S. K. Haq, and R. H. Khan. "Low Dose mercury Toxicity and Human Health." *Environmental Toxicology and Pharmacology* 20(2005):351 – 360.

Zahran, H. "Statistical Methods for Asthma Call Back Survey." Interview by author, 2011.

Zhu, L., B. P. Carlin, and A. E. Gelfand. "Hierarchical regression with misaligned spatial data: relating ambient ozone and pediatric asthma ER visits in Atlanta." *Environmetrics* 14(2003):537–557.