The Association Between Exposure to Traffic-Related Air Pollution During Pregnancy and Children's Health Outcomes in the San Joaquin Valley of California: An Example of Causal Inference Methods

By

Amy Michelle Padula

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Epidemiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Kathleen Mortimer, Co-Chair
Professor Ira Tager, Co-Chair
Professor Alan Hubbard
Professor Michael Jerrett

Spring 2010

Abstract

The Association Between Exposure to Traffic-Related Air Pollution During Pregnancy and Children's Health Outcomes in the San Joaquin Valley of California: An Example of Causal Inference Methods

by

Amy Michelle Padula

Doctor of Philosophy in Epidemiology

University of California, Berkeley

Professor Kathleen Mortimer, Sc.D., M.P.H., Co-Chair
Professor Ira Tager, M.D., M.P.H., Co-Chair

Ambient air pollution and traffic exposure are widely recognized as an important public health concern. This research aims to investigate the association between traffic-related air pollution exposure during pregnancy and two important public health outcomes: pulmonary function in asthmatic children and term low birth weight. Asthma is the leading cause of childhood morbidity and term low birth weight is an important predictor of infant mortality. The period of pregnancy may be a critical time during which exposures may affect these health outcomes.

Two study populations are used in this dissertation: the Fresno Asthmatic Children and Environment Study – Lifetime Exposure (FACES-LITE) and the Study of Air pollution, Genetics and the Early life events (SAGE). FACES-LITE is a longitudinal cohort of asthmatic children, aged 6-11 at baseline, with periodic pulmonary function tests and exposure assessment of ambient air pollutants during pregnancy in Fresno, California. SAGE is a study of birth records from four counties in the San Joaquin Valley of California from 2000-2006 linked to traffic density metrics based on the geo-coded residences of the mother at birth. For both studies, causal inference methods were used to estimate the association between exposure to traffic-related air pollution during pregnancy and these child health outcomes. Specifically, targeted maximum likelihood estimation (TMLE) was used to obtain the counterfactual marginal effect of traffic-related air pollution exposure during pregnancy on pulmonary function and term low birth weight. In other words, the predicted outcomes were compared had everyone been exposed to specific levels of air pollution during pregnancy.

The results of the TMLE for FACES-LITE found that above-median levels of ambient NO_2 exposure during the first and second trimesters were associated with deficits in pulmonary function for all age groups. The SAGE analysis showed the highest quartile of traffic density exposure was associated with significantly higher term low birth weight compared to the lowest quartile; however, there was no evidence of a monotonic exposure-response relation. In general, the studies presented in this dissertation suggest that traffic-

related air pollution exposure during pregnancy may be associated with pulmonary function deficits in children with asthma, as well as with an increased risk for term low birth weight.

These analyses represent the first application of TMLE to the study of air pollution and child health outcomes. In addition to their novelty, these causal inference methods are unique in that they offer easily interpretable parameters with important public health implications and unlike traditional regression methods, they do not assume arbitrary models. The analysis of the FACES-LITE study contributes to the subject-matter and supports earlier work on the association of ambient air pollution exposure during pregnancy and lung function in children by using the repeated measures of lung function. In contrast, the SAGE analysis focused on a methodological approach using causal methods and contextual variables. For that reason, I included only one exposure metric and one birth outcome for a demonstration of these methods. This subject-matter analysis will be extended in future analyses to further characterize the complexity of the exposure and any additional potential confounders and effect modifiers.

Acknowledgements

A special thanks to Kathleen Mortimer for being an inspiring mentor, advisor, and role model. It has been a pleasure working so closely with and learning from her throughout this process.

Thanks to Ira Tager for all of the guidance and training throughout my entire doctoral program.

Thanks to Alan Hubbard for all of his help with the statistical methods.

Thanks to Michael Jerrett for his advice with the traffic exposure assessment and his help facilitating the geocoding of the residences for the SAGE study.

Thanks to the funders, investigators and research staff of Fresno Asthmatic Children's Environment Study (FACES): Funded by NHLBI of NIH (R01 HL081521) and California EPA Air Resources Board (99-322) – Investigators and research staff: *Ira Tager, Katharine Hammond, John Balmes, Kathleen Mortimer, Romain Neugbauer, Jennifer Mann, Betsy Noth, Mark van der Laan, Helene Margolis, Frederick Lurmann, Siana Alcorn, Leah Cynthia Appel, Kathy Butler, Alex Gabaldon, Raul Gallegos, Leah Melendez, Mia Ortega, and Pamela Powers, Boriana Pratt, Beth MacDonald, Lucas Carlton, Meagan Loftin, Jessie Murphy, Melanie Gendell and the FACES participants and their families.*

Thanks to the funders, investigators and research staff of FACES Lifetime Exposures (FACES LITE): Funded by the American Lung Association, National and East Bay of CA Chapters – Investigators and research staff: *Kathleen Mortimer, John Balmes, Romain Neugebauer, Ira Tager, Frederick Lurmann, Siana Alcorn*

Thanks to the funders, investigators and research staff of the Study of Air pollution, Genetics and Early life events (SAGE): Funded by the NIEHS of NIH (1R21ES014891 – 01A2) – Investigators and research staff: *Kathleen Mortimer, Ira Tager, Elizabeth MacDonald, Frederick Lurmann, Siana Alcorn, Michelle Wilhelm-Turner, Toshiro Nishimura*

Thanks to Alice Pressman, Andy Avins, Michelle Odden, Jon Snowden and Sara Gale for all of your help with proofreading, code checking, census data compilation, lunch and chocolate breaks and moral support.

Thanks to my family: Mom, Dad, Anthony, Michael, Julia, Alice, Grace, and Clare for your love and support.

Thanks to Scott Rossi for your patience and companionship. And, thanks to the animals – DRP and Nutmeg – for chewing and sitting on my papers when I wasn't reading them.

Table of Contents

Introduction

Ambient air pollution and traffic exposure are widely recognized public health concerns. This dissertation research addresses the association between exposure to traffic-related air pollution during pregnancy and children's health outcomes in the San Joaquin Valley of California. With two separate study populations, I examine the effects of traffic-related air pollution on vulnerable populations – asthmatic children and pregnant women. The Fresno Asthmatic Children and Environment Study (FACES) is a cohort of asthmatic children (aged 6-11 at baseline) with periodic pulmonary function tests administered over 8 years of follow-up in Fresno, California. Additional exposure to ambient air pollution during the prenatal period was assessed in a sub-cohort, FACES Lifetime Exposure (FACES-LITE). The second study population, the Study Air pollution Genetics and the Early life events (SAGE), is a cohort of births between 2000-2006 that are linked to traffic density metrics in four counties in California (Fresno, Kern, Stanislaus and San Joaquin).

The motivation for this research is to investigate the association between traffic-related air pollution and important public health outcomes in children to inform regulatory guidelines. Studies such as these also are instructive for the design of communities and built environments. For example, the placement of schools and housing relative to high-traffic areas has been an important area of consideration. In addition, this dissertation is an application of causal inference methods to estimate easily interpretable parameters of interest at the population level, which are more informative about possible intervention effects of alterations to traffic exposure.

The first three chapters synthesize the background literature on the subject matter included in this dissertation. Chapter 1 summarizes the general health effects of ambient air pollution and traffic exposure and discusses the methodological issues surrounding the study of air pollution. This chapter also describes the standards for air quality and vehicle emissions in the United States and exposure assessment of traffic-related air pollution. Chapter 2 describes asthma and pulmonary function and summarizes the studies on associations between air pollution and pulmonary function in children. Chapter 3 defines adverse birth outcomes and addresses their public health importance. This chapter also reviews the epidemiological literature concerning the effects of exposure to air pollution during pregnancy and birth outcomes as well as the methodological issues encountered in these studies.

Chapter 4 describes the methods used in this dissertation. First, the two study populations and data acquisition procedures are described in more detail. Also included is an introduction to causal inference methods including the counterfactual framework, implementation of three estimators, and contrast to traditional regression methods. Chapter 5 presents the analysis of the FACES-LITE study, in which I use targeted maximum likelihood estimation to evaluate the association between exposure to ambient air pollution during pregnancy and repeated measures pulmonary function in childhood. Finally, Chapter 6 examines several methodologies to evaluate the association between exposure to traffic density during pregnancy and term low birth weight in the SAGE cohort.

Chapter 1: Health Effects of Ambient Air Pollution and Traffic Exposure

Exposure to ambient air pollution is recognized as an important health problem, both nationally and worldwide. Despite the passage of the Clean Air Act, the air in many parts of the United States, including parts of California, has concentrations of pollutants at which adverse health effects are observed. Certain populations are more susceptible to the harmful effects of air pollution including asthmatics, children and infants (Kim 2004), the elderly, and those with chronic obstructive pulmonary disease and underlying cardiovascular disease (Dominici, Peng et al. 2006). This chapter describes exposure assessment of traffic-related ambient air pollution and general health effects of air pollution. The effects of ambient air pollution on more specific child health outcomes will be discussed in Chapter 2 and 3.

1.1 Air Quality and Emission Standards

In 1952, the great London smog occurred when stagnant weather conditions coupled with a low-level thermal inversion trapped coal smoke in the Thames Valley and caused a sharp increase in the concentration of air pollutants. Over several days, more than three times as many people died than expected, leading to an estimated excess death toll of over 4000. It has been estimated that more than 12,000 additional deaths in the following months were due to these increases in air pollution. Conditions have changed since then and although much of the pollution, in general, is much lower than 50 years ago, certain components have gained prominence (Brunekreef and Holgate 2002).

In the 1960s, 1970s, and 1990s, the United States Congress enacted a series of Clean Air Act amendments which significantly strengthened regulation of ambient air pollution (USEPA 2010). The Clean Air Acts set numerical limits on the concentrations of lead and five air pollutants: nitrogen dioxide (NO_2), carbon monoxide (CO), particulate matter (PM), ozone (O_3) and sulfur dioxide (SO_2), known as "criteria" pollutants, and provided reporting and enforcement mechanisms. The introduction of catalytic converters began in 1975 to decrease the emitted pollutants from vehicle tailpipes and lead was phased out of gasoline from 1973 through 1995.

However, during the 1980s the number of motor vehicles in urban areas steadily increased and air quality problems associated with motor vehicles became more prevalent. In the early 1980s, the main interest was the effect of lead pollution on human health, but by the late 1980s and early 1990s, the effects of other motor vehicle primary pollutants and secondary pollutants became a major concern. Nitrogen oxides from transportation increased through the 1990s (U.S.EPA 2010) and have been subject to more investigation along with the other criteria air pollutants.

The San Joaquin Valley of California includes many of the counties in the country that are designated "nonattainment" for three to four of the criteria pollutants, that is, their air contains levels above those set by the Clean Air Act's National Ambient Air Quality Standards (NAAQS). In Appendix 2, a table presents the Ambient Air Quality Standards at the State, National, and International level. In general, the California standards are stricter than those on the national level. The World Health Organization (WHO) has set most

ambitious guidelines for NO_2, PM_{10}, $PM_{2.5}$, O_3, and SO_2. See Appendix 3 for a map of the counties designated "nonattainment" of the NAAQS as of August 2008.

Concerns about the health effects of traffic-related pollution and the likely continued growth of the vehicle population have stimulated discussion of tighter vehicular emission regulations. These regulations include higher emission standards for new vehicles, use of clean fuels, less polluting engines based on primary engine technology and tailpipe emission controls, inspection and maintenance of in-use vehicles and transportation planning. In the past 40 years, California has adopted more stringent emissions requirements for cars and light trucks.

1.2 Exposure Assessment of Ambient Air Pollution and Traffic Exposure

Assessment of the public health impact of traffic-related ambient air pollution exposure depends heavily upon estimates of air pollution concentrations. Estimation of exposure to ambient air pollution from traffic emission in epidemiological studies can be made from measurement of traffic-related pollutants and from direct measures of traffic itself. Assessing exposure of the population to traffic-related air pollution is complicated by several factors, including time activity patterns, meteorological conditions, land use pattern, topography and traffic conditions (HEI 2010).

1.2.1 Ambient Air Pollutants

Central air pollutant monitoring stations were established by the U.S. Environmental Protection Agency (U.S. EPA) to monitor criteria pollutants and measure population exposures on urban or regional scales. These fixed-site stations measure for different lengths and periods of time depending on the pollutant. For instance, carbon monoxide is generally measured every hour (in the San Joaquin Valley of California), but particle matter is usually measured every 6^{th} day. The measurements are then averaged over various periods. Annual air quality standards typically refer to averaging periods of 1, 8, and 24 hours (Appendix 2).

One limitation of pollutants measured at central monitors is the lack of spatial resolution necessary to capture both the temporal and spatial variability of pollutants from local-scale traffic. Concentrations may be lower in areas more distant from the monitors, which could cause bias in a study examining an association with a health effect (HEI 2010). Exposure misclassification varies by averaging period of the pollutant and the length of exposure period. Additionally, an exposure period of a trimester versus the entire pregnancy will affect the variability of the measurement. Factors that may help characterize the accuracy and precision of these estimates include time activity patterns and meteorological conditions.

Nitrogen dioxide (NO_2) is emitted from high temperature combustion from heating sources, power generation and motor vehicles. Emissions of NO_x (nitrogen oxides) are increased with higher vehicle activity, particularly diesel engines, though lower emission technologies are more widely used than in the past. On-road vehicles account for 33% of NO_x (U.S.EPA 2008; HEI 2010), primarily in the form of NO. Much of the NO_2 arises from the oxidation of nitrous oxide (NO) by oxygen (O_2) in the presence of sunlight and decays exponentially with distance from traffic. In many outdoor environments, nitrogen dioxide concentrations are primarily related to traffic-related combustion products, notably particle

matter (Brunekreef and Holgate 2002). Average national NO_2 ambient concentration in 2007 decreased 43% from 1980 (U.S.EPA 2008).

Carbon monoxide (CO) is colorless, odorless, non-irritating but very poisonous gas. It is a product of incomplete combustion of fuel such as natural gas, diesel and gasoline fuels, coal or wood. Vehicular exhaust is a major source of anthropogenic CO; on-road vehicles account for approximately 47% and off-road vehicles account for 21% of total CO emissions (U.S.EPA 2008; HEI 2010). Concentrations of CO have shown a downward trend over the past decade in California as well as the entire U.S. due in part to emission controls and catalytic converters of on-road vehicles(Holgate 1999).

Particulate matter (PM) is not a single pollutant, but is made up of particles of different sizes and chemical composition. PM can consist of different particles varying with space and time due to changes in traffic volume, distance from roads, topography, and meteorology. PM arise from motor vehicles and as by-products of reactions that form secondary particles (HEI 2010). PM_{10} is the fraction of suspended particles that pass an inlet with a 50% cut-off efficiency at an aerodynamic diameter of 10 μm. $PM_{2.5}$ is particle matter less than or equal to 2.5 μm in aerodynamic diameter and comes from combustion and conversion of gases to particles. According to a recent study (Watson, Chen et al. 2008), approximately 15% of $PM_{2.5}$ is attributable to motor-vehicle emissions (HEI 2010). Prior to these definitions, PM was measured as total suspended particles (TSP) or Black Smoke (BS). PM_{10} was adopted by the U.S. EPA in 1987 and the distinction of $PM_{2.5}$ was added in 1997. Diesel vehicles are significant sources of PM in California and the U.S. Gasoline-fueled cars have decreased the amount of PM and increased oxides of nitrogen and sulfur emitted, but the large and increasing number of vehicles on the road maintains its contribution of PM to the air. Due to the traffic-related sources of $PM_{2.5}$ and NO_2, these two pollutants are often highly correlated (Sarnat, Schwartz et al. 2001). As mentioned above, the separation of these effects can be difficult when the causal effects of individual pollutants are of interest.

Ground level ozone (O_3) forms in the atmosphere from the reactions in sunlight between NO_x and volatile organic compounds (VOCs), which result in higher O_3 in summer months. It is not a pollutant emitted from motor vehicles as the pollutants mentioned above, however, it is included because in interacts with traffic-related pollutants and is an irritating and regulated pollutant. The correlation of O_3 and other pollutant concentrations in outdoor air is often low in California, so the effects of O_3 and other pollutants can be separated relatively easily. In addition, wind flows can move ozone-forming air masses over many hundreds of miles. Ozone can have adverse effects not only on human health, but also on crops, vegetation and ecosystems, as well as on materials.

Sulfur dioxide (SO_2) is emitted from burning of coal and oil. Emissions of SO_2 have decreased in the past few decades due to transition from soft to hard coal to lower sulfur-containing fuels. Although road transport is now a minor source of SO_2, in some urban areas, higher concentrations have been detected along busy roads. In California, sulfur oxides (SO_x) are not as high as it is in some other parts of the U.S., particularly the East Coast where sulfuric acid has been transported from power plants in the Midwest (Holgate 1999). Although SO_2 is one of the criteria pollutants, it is not under investigation in this dissertation because of the lesser role it has in California's ambient air pollution.

1.2.2 Traffic Exposure

Motor vehicles are a major source of ambient air pollution, particularly in California. A major fraction of the global population spends significant time on or near roadways as a part of their daily activities. Although progress has been made in reducing emissions from individual vehicles, there has been substantial growth in the number of vehicles and miles traveled in the U.S. in the past 15 years (HEI 2010). U.S. travel data indicate commuters spent 81 minutes per day in vehicles in 2001, on average, which was 10% higher than in 1995, and children spent 48 minutes per day in vehicles (Hu 2004). Motor vehicles contribute to CO, $PM_{2.5}$, NO_x and other pollutants to the air and concentrations of these traffic-related pollutants are greater near major roads (Zhu, Hinds et al. 2002). Diesel combustion, in particular, is a major source of NO_2 and PM. At present only a small fraction of new cars in the U.S. have diesel engines, but industry analysts expect this fraction to grow significantly in the future as the technology lower the emissions from new diesel engines enough to meet California's strict emissions standards.

There have been many advances in exposure modeling for traffic-related air pollution over the past decade, in part due to the increased availability of geographical information system (GIS) and associated modeling techniques (Briggs 2007). The most basic exposure assessment methodologies are proximity-based where indicators of the relative concentrations of vehicle-related pollution are used to estimate individual traffic exposure (HEI 2010). One commonly used metric is distance to nearest roads. Experimental studies and dispersion theory indicate that pollution levels generally decrease inversely with distance from roadways. Thus, distance (or inverse distance) to various types of roadways is a potential indicator of exposure to traffic-related pollution. These distances are often specified by road classes 1-4 and assigned respectively to freeways, primary highways, secondary highways, and local roads. An additional traffic metric frequently included in the published literature is length of roads in a 200 meter radius buffer around each location of interest. This metric is also specified by road class (1-4). This proximity metric may represent the density of roads in an area better than the distance to the nearest road. Proximity models usually provide a relatively crude but easy to implement evaluation of the impact of traffic pollution on health. These models are limited to the statistical investigation of traffic activity in relation to the risk of respiratory illness (HEI 2010). There is concern about the accuracy of proximity models and its value as a surrogate for exposure assessment to traffic-related air pollution.

An additional metric that will be used in Chapter 6 is a dimensionless indicator of traffic density based on distance-decayed annual average daily traffic (AADT) volumes. The traffic density is calculated with roadway link-based traffic volumes which are derived from traffic count data. An advantage to the traffic density parameter is that it accounts for the combined influence of all roadways and activity (for which data exist) near each location (Penfold 2009). Two main problems are associated with the resolution and accuracy of traffic count data. The first is the tendency for agencies to count traffic more thoroughly in more heavily trafficked areas, leading to missing data and undercounts in light-traffic areas. The second problem results from data sets containing unrepresentative counts that might span long periods and traffic-count bias caused by the time of year in which the short-term counts were made (HEI 2010).

Jerrett *et al.,* have described additional modeling approaches for traffic exposure-assessment in more detail (Jerrett, Arain et al. 2005). In addition to the metrics discussed above, studies of health effects use additional the following techniques: geostatistical interpolation, land use regression (LUR) models, dispersion models, and hybrid models that combine these approaches with time activity or personal monitoring data used to derive some measure of individual exposure (Jerrett, Arain et al. 2005). Interpolation models estimate pollution concentrations at across space between measuring sites. Factors are incorporated such as topography, local emissions and variability of the measured pollutant. LUR models treat the pollutant of interest as the dependent variable and proximate land-use, traffic, and physical environmental variables as independent predictors. The predict pollution concentrations at a given site based on surrounding land use and traffic characteristics. Dispersion models incorporate meteorological data. These models are potentially more reliable than the models described above if meteorological data are precise (HEI 2010). These models are limited to the degree of temporal and spatial resolution of the data.

Personal monitors can be used on study subjects to measure personal exposure directly. This option accounts for time spent at various locations and incorporates the estimation of indoor air pollution, though they are not feasible for estimation of exposure for large populations or over extended periods of time nor are they practical for active or young populations. Additionally, they do not have the ability to separate ambient from indoor air pollution without time-activity data. For this reason, personal monitoring is often used as a complement to other model types to create hybrid models.

In summary, accurately measuring ambient air pollutants and traffic exposure in both time and space is difficult in large health studies. The methods mentioned above have been used in studies of health effects and the increase in technological innovation, GIS and geostatistical advances continues to improve these methods. The next section gives a summary of the range of health effects attributed to ambient air pollution and traffic exposure.

1.3 Health Effects of Air Pollution and Traffic Exposure

Current scientific understanding of the spectrum of health effects related to exposure to ambient air pollution has increased substantially in the past two decades, and numerous studies have found important health effects from air pollution at levels once considered safe. Ambient air pollution has been associated with numerous adverse health outcomes including acute effects such as respiratory symptoms, asthma exacerbations, decreased pulmonary function, cardiovascular events, and hospital admissions as well as long-term effects such as chronic bronchitis, markers of atherosclerosis, and cardiovascular mortality (Kim 2004; HEI 2010).

Human health concerns of NO_x include effects on breathing and the respiratory system, damage to lung tissue, and premature death. Small particles penetrate deeply into sensitive parts of the lungs and can cause or worsen respiratory disease such as emphysema and bronchitis, and aggravate existing cardiovascular disease (Pope and Dockery 2006).

Carbon monoxide combines with hemoglobin to form carboxyhemoglobin and hinders the delivery of oxygen to the body (Kim 2004). High CO concentrations have been associated with early onset of cardiovascular disease, behavioral impairment, decreased

exercise performance, sudden infant death syndrome (SIDS) and increased daily mortality rates (Fierro 2000).

Particle pollution contains microscopic solids or liquid droplets that are so small that they can get deep into the lungs and cause serious health problems. Numerous scientific studies have linked particle pollution exposure to a variety of problems, including increased respiratory symptoms, such as irritation of the airways, coughing, or difficulty breathing; decreased lung function; aggravated asthma (Kunzli and Tager 2000); development of chronic bronchitis (Kunzli and Tager 2000); irregular heartbeat; nonfatal myocardial infarctions; and premature death in people with heart or lung disease.

Ozone can irritate the respiratory system and inflame the lining of the lung which can lead to permanent changes in the lung. Ozone has been associated with respiratory symptoms, reduced pulmonary function and airway inflammation (Mudway and Kelly 2000; Tager, Balmes et al. 2005).

Sulfur dioxide causes a wide variety of health and environmental impacts because of the way it reacts with other chemicals to form sulfate particles. Peak levels of SO_2 have been associated with increased respiratory symptoms and disease, difficulty in breathing, and premature death. Although SO_2 itself largely does not get past the nose, effects are likely related to sulfuric acid (H_2SO_4) formed in the atmosphere (Tewari and Shukla 1991).

Diesel exhaust is carcinogenic and diesel exhaust particles (DEP) increase airway inflammation and can exacerbate and initiate asthma and allergy (Bernstein, Alexis et al. 2004). Increased adverse health effects were found among those living near busy roads (Delfino 2002). Increased respiratory tract complications have been associated with residences near areas of high traffic density (Brunekreef, Janssen et al. 1997).

1.3.1 Previous Studies

Early ecological studies demonstrated increased respiratory mortality rates in higher pollution regions (Collins, Kasap et al. 1971; Lave and Lave 1977; Barker and Osmond 1986; Bobak and Leon 1992). These studies, however, lack information on individual characteristics and it is unknown how other characteristics, such as co-pollutants, meteorology, smoking prevalence, age, race and physical activity levels that differ between regions may have confounded the results.

Time series studies have shown changes in mortality and hospital admissions due to day-to-day variation in air pollution (Bates and Sizto 1987; Pope, Dockery et al. 1991; Schwartz 1991; Burnett, Dales et al. 1994; Thurston, Ito et al. 1994). Advantages of these studies include contrasts in exposure due to weather-driven variation over time and populations serve as their own controls. However, they also are not well suited to investigations of individual-level factors and assume the confounding by trends is sufficiently controlled. It is possible that the contrast in air pollutant levels over time does not vary to a large enough degree to detect an effect.

Long-term studies have evaluated the effects on cohorts of individuals exposed to air pollutants over time with the ability to condition on individual characteristics such as smoking. Three studies of adults in the late 1970s through the 1980s – Harvard Six Cities Study, American Cancer Society (ACS) Study and Adventist Study of Smog (ASOSMOG) -

suggested exposure to fine particulate matter in the air was associated with higher mortality, particularly cardiopulmonary and lung cancer mortality (Abbey, Mills et al. 1991; Dockery 1993; Pope, Thun et al. 1995). Through further analyses, investigators found that the daily increases in mortality with increases in daily PM were not simply due to harvesting (Zeger, Dominici et al. 1999; Schwartz and Neas 2000; Schwartz 2001).

In the Six Cities study, after adjustment for smoking and other risk factors, Dockery *et al.* observed associations between air pollution and mortality. The adjusted mortality-rate ratio for the most polluted of the cities as compared with the least polluted was 1.26 (95 % CI: 1.08-1.47). Air pollution was positively associated with death from lung cancer and cardiopulmonary disease but not with death from other causes considered together. Mortality was most strongly associated with air pollution with fine particulates, including sulfates (Dockery 1993).

The ACS study collected data as part of the Cancer Prevention II Study in 1982. The risk factor data for approximately 500,000 adults were linked with air pollution data for metropolitan areas throughout the United States and combined with vital status and cause of death data through 1998. Pope *et al.* found that fine particulate and sulfur oxide–related pollution were associated with all-cause, lung cancer, and cardiopulmonary mortality. Each 10-$\mu g/m^3$ elevation in fine particulate air pollution was associated with approximately a 4%, 6%, and 8% increased risk of all-cause, cardiopulmonary, and lung cancer mortality, respectively (Pope, Burnett et al. 2004).

The AHSMOG Study was a longitudinal study of 6,000 residentially stable and non-smoking Seventh-day Adventists in California conducted to evaluate long-term cumulative effects of ambient air pollution on several chronic diseases. It was one of the first prospective studies to observe an association between long-term exposures to ambient air pollutant concentrations on the occurrence of incident asthma. Multivariate analyses which adjusted for past and passive smoking, and occupational exposures, indicated statistically significantly elevated relative risks ranging up to 1.7 for incidence of asthma, definite symptoms of airway obstructive disease, and chronic bronchitis with TSP in excess of all thresholds (100 $\mu g/m^3$, 150 $\mu g/m^3$ and 200 $\mu g/m^3$), corresponding to California and national standards, except the lowest one (60 $\mu g/m^3$) (Abbey, Mills et al. 1991).

More recent studies have begun to look at traffic exposure in addition to individual pollutants. Many studies have reported health effects associated with residential proximity to traffic and traffic density. In a review of 29 studies (Boothe and Shendell 2008), 25 reported statistically significant associations with at least one adverse health effect over a broad range of exposure metrics and diverse geographical locations. Uncertainties exist because of the inability to control for confounding in many of these studies.

Studies have investigated mortality (both all-cause and cardiovascular), cardiovascular morbidity, asthma incidence and exacerbation, respiratory symptoms (such as cough, phlegm and wheeze), lung function, health care utilization for respiratory disease, allergy, birth outcomes, and cancer. For a more detailed review of these studies, I refer you to the Health Effects Institute Monograph on Traffic-Related Air Pollution (HEI 2010). This critical review synthesizes the literature and the committee judges the evidence for causal associations for traffic-related exposures and health outcomes. They concluded "sufficient" evidence to

support a causal association between traffic-related air pollution and exacerbation of asthma and "suggestive" evidence of a causal association with onset of childhood asthma, non-asthma respiratory symptoms, impaired lung function, all-cause and cardiovascular mortality, and cardiovascular morbidity (HEI 2010).

To highlight a few studies, in subset of a Dutch cohort, Hoek *et al.* (Hoek, Brunekreef et al. 2002) found evidence of increased cardiopulmonary mortality among those living near a major road (within 100 m of a highway or 50 m of a major road). However, a more recent follow-up with a larger sample demonstrated smaller risk estimates (Beelen, Hoek et al. 2008).

Another study in the U.S. found traffic density within a 100 meter buffer, distance from roadway controlling for socioeconomic factors associated with acute myocardial infarction (Tonne, Melly et al. 2007). Other investigators have linked various childhood cancers to proximity to traffic in Denver, Colorado (Savitz and Feingold 1989; Pearson, Wachtel et al. 2000).

Certain populations are more vulnerable to the effects of traffic-related air pollution. People with diseases such as asthma and cardiovascular disease and people who work or exercise outside are susceptible to adverse effects such as damage to lung tissue and reduction in lung function.

Infants and children are among the most susceptible to many of the air pollutants because of their developing immune system and the lung growth and development that occurs throughout gestation and childhood. Epidemiologic data have established associations between prenatal exposure to ambient air pollutants and a variety of adverse birth outcomes including intrauterine mortality (Pereira, Loomis et al. 1998), low birth weight (Ritz and Yu 1999.; Wang, Ding et al. 1997), preterm birth (Ritz, Yu et al. 2000), small for gestational age (Dejmek, Solansky et al. 2000; Liu, Krewski et al. 2003), and neonatal mortality (Loomis, Castillejos et al. 1999), and postnatal mortality (Woodruff, Grillo et al. 1997; Bobak and Leon 1999).

The consequences of low birth weight (LBW) and preterm birth (PTB), in particular, have been studied extensively and are associated with considerable short- and long-term health effects. Among those health effects, LBW and PTB have been associated with asthma in childhood. It has been suggested that the fetal period and early childhood play an important roles also for asthma and other allergic diseases (Bjorksten 1999). It is possible the growth of the lung is altered due to these fetal exposures. Chapter 3 will review the literature on prenatal effects of air pollution on birth outcomes more thoroughly.

A recent review concluded that there is strong evidentiary support for an adverse effect of air pollution on lung function in children and adolescents (Gotschi, Heinrich et al. 2008). The variety of study designs approaches to exposure assessment, and lung function measures that have been used to study this question make it difficult to synthesize the results. Many studies have been done on the short term effects of ambient air pollution on lung function; however, the more important public health concern is the potential long term effects of long time exposure to ambient air pollution, particularly for children. This vulnerability is also highlighted by the fact that children tend to spend more time outside than adults, increasing their exposure further to ambient air pollutants (Dietert, Etzel et al. 2000; Pinkerton 2007). Data do suggest that long term exposure to air pollution is associated with deficits in lung

function growth (Gauderman, Vora et al. 2007) and deficits in lung function have been shown to be a strong determinant of life expectancy (Hole, Watt et al. 1996). A more thorough review of the literature on the effects of air pollution on pulmonary function in children is in Chapter 2.

In addition to epidemiological studies, controlled human exposure studies have investigated the effects of air pollution on health. Although these studies often have small sample sizes, short durations of exposure and simple environments, some important findings have been suggested. These studies have found evidence of acute inflammatory response (Holgate; Salvi, Nordenhall et al. 2000; Holgate and Peters-Golden 2003), decrements in lung function (McCreanor, Cullinan et al. 2007) and allergic responses in previously sensitized individuals (Strand, Rak et al. 1997). These studies have had a greater impact on the study of the mechanism by which ambient air pollution and traffic exposure may cause health effects, particularly in the lung.

One hypothesis is that oxidative stress (Sies 1991) is responsible for the adverse effects of ambient air pollution on human health. Traffic-related air pollution, in particular, can affect pulmonary health through reactions with the lining fluid of the pulmonary airways (Postlethwait, Langford et al. 1995; Pryor, Squadrito et al. 1995; Mudway and Kelly 2000) and, in turn, can cause inflammation and DNA damage. A more detailed discussion of the mechanism by which air pollution is hypothesized to cause human health effects is in Chapter 2.

1.3.2 Methodological Issues

This dissertation will only address ambient air pollution; however, indoor air pollution is also of concern for public health. Its sources include cooking and heating appliances, secondhand exposure to tobacco smoke (SHS), molds, household cleaning products, building materials and others.

Effects of ambient air pollution have been detected at very low levels of exposure, but it is not clear whether a threshold concentration exists for given pollutants below which no effects on health are likely. Although this would be useful for public health efforts, it is difficult to determine. Epidemiological issues such as misclassification of exposure, combination of exposure (mixtures of pollutants), correlated exposures and the absence of a large contrast of exposure are all challenges to this question. For instance, virtually no one is "unexposed" to ambient air pollution. Everyone is exposed to various pollutants at various levels for cumulatively over time, and therefore it is impossible to compare to an unexposed control. Although often studied and analyzed separately, it is likely that it is the total pollutants rather than any single component that is responsible for these associations.

Many of the studies mentioned above used central monitors to assess exposure. As mentioned above (1.2.1), a limitation of this exposure assessment is the lack of spatial resolution. This resolution is necessary to capture the spatial variability of pollutants on a local scale. The traffic exposure studies have used more sophisticated modeling techniques to assess traffic density, volume and proximity to estimate exposure of populations to traffic-related air pollution. The limitations of these methods are also mentioned above (1.2.2).

Additional problems exist for studies on air pollution and health effects. For cross-sectional studies, establishment of a temporal sequence may be difficult. There may be a lack

of data on the duration of residence in the area studied. Survivor effects due to out-migration or early death from specific disease may cause bias in a study. Age, cohort and period effects can confound the relationship between air pollution and disease. Finally, the effects of air pollution may be small on an absolute scale and therefore difficult to detect, however, because so much of the population is exposed to air pollution, the public health significance is enormous.

In summary, this chapter has introduced the public health importance of exposure of ambient air pollution and one of its most important sources, traffic-related air pollution. The methods for measuring and assessing exposure and the methodological issues surrounding this field were introduced. As discussed, ambient air pollution exposure is associated with a range of health outcomes and the next chapter will discuss one in more detail, the effect on pulmonary function, particularly among children.

Chapter 2: Asthma and Pulmonary Function in Children

2.1 Asthma

Asthma is a chronic pulmonary disease characterized by air flow obstruction due to chronic inflammation of airways of the lung and episodic increases in airway resistance (bronchospasms) that are experienced by recurring periods of wheeze, chest tightness, shortness of breath, and cough. Longer term effects such as airway remodeling (permanent alterations in the airway structure) and fixed obstruction can be a consequence of asthma (Holgate and Polosa 2006).

Airway inflammation involves an interaction of many cell types and multiple mediators with the airways that eventually results in the pathophysiological features of asthma: bronchial inflammation and airway narrowing that result in recurrent episodes of cough, wheeze and shortness of breath. The processes by which these interactive events occur and lead to asthma are still under investigation. It is well established that asthma is a variable disease. Distinct phenotypes of asthma exist (*e.g.,* intermittent, persistent, exercise-induced, severe), however, airway inflammation remains a consistent pattern. Furthermore, the natural history of asthma varies in different age groups (ERP 2007). This is discussed further in section 2.3.

Asthma is the most important chronic disease of childhood in terms of numbers affected, morbidity and health care costs (Wang, Zhong et al. 2005). The prevalence of asthma, particularly among children, has escalated over the past three decades, and has resulted in an increase in asthma hospitalization rates. There are currently 6.7 million (9.1%) children under 18 years old with asthma in the U.S. (Bloom, Cohen et al. 2009). An estimated 11.9% of Californians – 3.9 million children and adults – report that they have been diagnosed with asthma at some point in their lives, compared to the national average of 10.1%. Nearly 667,000 school-aged children in California have experienced asthma symptoms during the past 12 months (Weller 2010).

There is extensive evidence that asthma is exacerbated by exposure to traffic-related pollutants although fewer studies show evidence that pollutants cause asthma (Braback and Forsberg 2009). The burden of disease is not shared equally. Neighborhoods characterized by a higher percentage of minorities, lower incomes, inadequate housing, and ambient air pollution are correlated positively with asthma hospitalization rates (Corburn, Osleeb et al. 2006). Increased asthma prevalence and asthma hospitalizations have been associated with levels of deprivation in New Zealand and England (Watson, Cowen et al. 1996; Salmond, Crampton et al. 1999).

Asthma is diagnosed with a clinical examination including medical and symptom history and spirometry, which is discussed in the following section. Spirometry measures obtained from forced expiration maneuvers are used to measure pulmonary function and assess asthma severity. Further below (Section 2.4), I will discuss previous research on the effects of air pollution on asthma and measures of pulmonary function.

2.2 Spirometry

Spirometry is the most common kind of pulmonary function test (PFT) to measure lung function. In addition to its use in the diagnosis and progression of asthma, levels of lung function are predictors of future morbidity and mortality (Sorlie, Kannel et al. 1989). It measures the amount (volume) and/or speed (flow) of air that can be inhaled and exhaled. Many factors influence the results of these tests including age, gender, body height and size, health status, and race. Abnormalities in PFTs can be attributed to asthma and other pulmonary diseases. A series of PFTs are listed below, some of which are particularly useful to evaluate asthma severity. All of the following PFTs are used in the analysis in Chapter 5.

Description of pulmonary function tests

Spirometry Measure	Abbreviation	Description
Forced vital capacity	FVC	Amount of air forcibly expelled after full inspiration, measured in liters
Forced expiratory volume in one second	FEV_1	Amount of air forcibly expelled in one second, measured in liters
Forced expiratory flow 25-75%	FEF_{25-75}	Average speed of air expelled in the middle portion of the expiration (between 25% and 75% of vital capacity)
Forced expiratory flow at 75%	FEF_{75}	Amount of air forcibly expelled at 75% of vital capacity
FEV_1/FVC ratio	FEV_1/FVC	Ratio of forced expiratory volume in one second to forced vital capacity
FEF_{25-75}/FVC ratio	FEF_{25-75}/FVC	Ratio of forced expiratory flow 25-75% to forced vital capacity

FEV_1, most widely used measurement of large airways in epidemiologic studies, is found to be lower in those with obstructive disease such as asthma. However, the small airways are the principal sites of chronic obstruction in asthma. FEF_{75} is used as a measure of small airway function. FEF_{25-75} may be a more sensitive marker of small airway obstruction, though its reproducibility is poor. FEF_{25-75}/FVC is the ratio of forced expiratory flow 25-75% to forced vital capacity, which has the interpretation of the reciprocal of the time constant of the lung (Tager 1986), similar to Meade's $Vmax_{50}/(VC \times Pst(L)_{50})$ (*i.e.,* instantaneous flow at 50%, divided by vital capacity times elastic recoil pressure at 50% of vital capacity) and is reflective of intrinsic airway size (Mead 1980). Deficits in pulmonary function tests are often suggestive of asthma exacerbations and are sometimes used as markers of asthma severity.

2.3 Pulmonary Growth in Children

A child's lung is not simply a miniature version of an adult lung. Lung growth during childhood involves both an increase in size of the individual components and extensive remodeling, meaning parts change shape and function (Stick 2000). Development of the lungs spans from embryogenesis to adult life, passing through several distinct stages of growth (Pinkerton and Joad 2006).

An important consideration in the determination of the effects of various environmental exposures on respiratory health in children is the state of development of the lungs and the immune system at the time of exposure. The lungs begin to develop at 6 weeks of gestation and continue through distinct phases of progression during the first and second trimesters. The airways and blood vessels are in place by 26 weeks' gestation (third trimester). The proliferation of alveoli occurs primarily from 26 weeks though birth, and continues into childhood. It is biologically plausible that air pollution exposures during this development and progression of lung growth can cause damage, however, it is inconclusive as to which period during this development is the most critical or if it may be a cumulative effect over time during the pregnancy or throughout one's life. Children have a larger lung surface area relative to their body weight than adults and, under normal conditions, breathe 50% more air relative to their body weight than adults (Schwartz 2004). In addition, children's immune systems continue to develop and children tend to spend more time outdoors than adults, exposing them to more ambient air pollution.

Environmental exposures during specific periods or windows of development may have profound effects that would not be seen if the same exposure were to occur in the adult. Studies of second hand smoke exposure *in utero* and in early life is associated with decreases in lung function, particularly in the small airways (Tager 2008). Cunningham *et al.* studied 9-11 year old children in Philadelphia and showed that *in utero* exposure to tobacco smoke had an effect on lung function after adjustment for postnatal smoke exposure with a 5% reduction in FEF_{25-75} and a 1.2% (NS) reduction in FEV_1/FVC (Cunningham, Dockery et al. 1994). A longitudinal study of 8706 children demonstrated that after adjustment for current maternal smoking, exposure to maternal smoking in the first 5 years of life was associated with significant decreases in FEV_1/FVC and FEF_{25-75} growth (Wang, Wypij et al. 1994).

Assessment of the growth of the lung has often focused on the change in pulmonary function in growing children (Sherrill 1990). Lung growth has two major components: maturation of the lung and the direct relationship with body mass or size. The growth of children's lungs has differential rates as does children's physical growth. Growth spurts occur near puberty and post-pubertal adolescence. Lung growth can be examined by change in spirometry measures or growth velocity (Sherrill, Holberg et al. 1990).

Decreased levels of lung function and declines in lung function growth observed in children appear to occur by 6 years of age and occur predominantly in those children whose asthma symptoms started before 3 years of age. Children 5–12 years of age who have mild or moderate persistent asthma, on average, do not appear to experience declines in lung function through 11–17 years of age, although a subset of these children experience progressive reductions in lung growth as measured by FEV_1. Furthermore, there is emerging evidence of reductions in the FEV_1/FVC ratio, apparent in young children who have mild or moderate asthma compared to children who do not have asthma, that increase with age (ERP 2007).

2.4 Risk Factors for Asthma

The phenotypic expression of asthma is a complex, interactive process that depends on two major factors: an individual's susceptibility and environmental exposures. Individual factors associated with asthma include the following: immunity, genetics, sex, low birth

weight/preterm birth, race/ethnicity, parental history of asthma, respiratory infections and experience in delays of receiving asthma care. Factors which decrease asthma risk include day care attendance, breast feeding history, health insurance, income, and asthma medication use.

Immune system responses are associated with the development and regulation of inflammation. An increasing number of studies address the genetic components to asthma and polymorphisms in inflammatory genes are associated with risk of asthma. More will be discussed below about the gene-environment interactions that have been identified with respect to air pollution. There are sex differences in the prevalence of asthma. In early life, the prevalence is higher in boys, though after puberty, asthma is more common in women. It is not clear the role of hormones in the onset and persistence of asthma. Infants born low birth weight or preterm have a higher incidence of asthma, than those born of normal weight and gestation (Steffensen, Sorensen et al. 2000; Halterman, Lynch et al. 2009). These associations may be due to the development of the lung and immune system *in utero.* Race is associated with asthma; African-Americans are among those with the highest prevalence of asthma in the U.S. (Schwartz, Gold et al. 1990; Bloom, Cohen et al. 2009). The number of respiratory viruses in infancy has been associated with asthma, though it is possibly this association is an indirect effect of atopy.

The factors listed above are all potential confounders and/or effect modifiers in the relationship between ambient air pollution and lung function. For instance, African-Americans experience higher levels of air pollution and a higher incidence of asthma in many parts of the U.S. (Meng, Wilhelm et al. 2007; Meng, Wilhelm et al. 2008). In addition to these personal factors, investigations have established "contextual variables" at the community level, which also affect pulmonary function and mediate the effect of air pollution on pulmonary function (Jerrett, Burnett et al. 2005). Evidence from social epidemiology suggests that neighborhood context may affect health independently beyond individual risk factors (Diez Roux 2001).

Neighborhood-level measurements, often obtained from the U.S. census, include socioeconomic position, median household income, proportion of respondents with low education, percent of males unemployed and percent living in poverty. Socioeconomic position, for example, has been associated with increased exposure to traffic related air pollution and to deficits in pulmonary function and should be considered as a potential confounder. Neighborhood socioeconomic position has also been examined as an effect modifier though results have been mixed (Wheeler and Ben-Shlomo 2005). Further studies are needed to resolve these conflicting results.

There is large variation between individuals in their response to air pollutants. There are genetic factors that influence the mechanisms of lung injury caused by air pollutants. Polymorphisms in oxidative stress and inflammatory genes influence the response to air pollutants and modify its effects on respiratory systems, pulmonary function, and risk of asthma (Yang, Fong et al. 2008). Identification of genes that influence the air pollution to asthma pathways can help to understand individual and population heterogeneity in response to ambient air pollution. Genetic subgroups that are differentially affected by air pollutants have begun to be identified. Glutathione S-transferase (GST) deletion genotypes were shown to play an role in susceptibility to the effects of oxidant pollutants such as diesel exhaust

particles (Adams and Eschenbach 2004). Studies have found that children with GST gene polymorphisms and exposure to ambient air pollution had a higher risk for asthma (Wang, Zuckerman et al. 2002), acute decrements in lung function in those who were asthmatic (Wang, Zuckerman et al. 2002) and a lower rate of respiratory infections (Nukui T, Day R et al. 2004; McConnell, Berhane et al. 2003). These studies provide strong evidence for a role of this GST gene polymorphism as one genetic determinant of the response to oxidative stress in airway cells and thus susceptibility to inhaled oxidant-induced toxicity.

Oxidative stress occurs when an excess of free radicals exceeds the available antioxidant defenses and increases cellular concentrations of oxidized lipids, proteins, and DNA. Traffic-related air pollution presents free radicals causing oxidative stress when metals (Mudway, Stenfors et al. 2004) and polycyclic aromatic hydrocarbons (Squadrito, Cueto et al. 2001) from the PM of exhaust enter the lung (Kelly, Wagner et al. 2003). Increased oxidant burden may also play a role in transcriptional activation of pro-inflammatory genes, contributing to tissue injury (Brauner, Forchhammer et al. 2007).

Environmental exposures such as allergens, tobacco smoke and air pollution can be on the individual or community level. Sensitization and exposure to allergens including pets, mold and cockroaches have play an important (yet, unclear) role in the development of asthma. Exposures to allergens have been associated with protecting against (Riedler, Braun-Fahrlander et al. 2001) and exacerbation of asthma (Huss, Naumann et al. 2001; ERP 2007). Tobacco smoke, as mentioned above, has been associated with the onset and exacerbation of asthma. The role of ambient air pollution in the development of asthma continues to be studied. The Children's Health Study found that exercise outdoors in communities with high ozone concentration was associated with a higher risk of asthma among school-age children (McConnell, Berhane et al. 2002). Section 2.5 (below) will address what is known about the effects of air pollution on asthma and lung function.

2.5 Previous Studies on the Effects of Ambient Air Pollution on Pulmonary Function in Children

Gotschi reviewed a diverse group of studies on the long-term effects of ambient air pollution on lung function in both children and adults. Synthesis of these data are complicated by the variety of ambient air pollutants and exposure assessment options, the multiple lung function measurements and the covariates measured in each study. Despite this heterogeneity of study design, support is strong for the hypothesis that there are adverse long-term effects of air pollution on pulmonary function growth in children, which result in deficits of pulmonary function at the end of adolescence (Gotschi, Heinrich et al. 2008). Air pollution measurements were predominantly made at the community level in various cities across the U.S. and world, and used centrally located monitors that sample the criteria pollutants at various time intervals (*e.g.,* 8-hour maximum, 10am-6pm average, 24-hour average) and a few included traffic exposures. The studies used spirometry to measure lung function (predominately FEV_1 and FVC) (Gotschi, Heinrich et al. 2008). Gotschi examined cross-sectional and longitudinal studies, most of which reported statistically significant adverse effects of air pollution on pulmonary function (Gotschi, Heinrich et al. 2008).

The most relevant (because of its size, study population and location) is the Children's Health Study (CHS) in southern California, which started in 1993. In the CHS, 1759 children ages 10-18 years in 12 communities have been followed longitudinally. Investigators found reduced lung growth in children exposed to higher levels of air pollution, both regionally and in proximity to traffic (Gauderman, McConnell et al. 2000; Gauderman, Gilliland et al. 2002; Gauderman, Avol et al. 2005; Gauderman, Vora et al. 2007). Children in the most polluted communities experienced growth deficits in FEV_1 of 100mL, corresponding to an approximate 7% decrease in girls and 4% decrease in boys, compared to those living in the least polluted communities. More specifically, in the CHS, lung function growth was 10% slower in communities with higher NO_2 levels (Kunzli, McConnell et al. 2003). The proportion of children with clinically low pulmonary function at age 18 ($FEV_1 < 80\%$ of predicted value) was estimated to be 5 times larger in the most polluted community (annual average $PM_{2.5}$ – 29 $\mu g/m^3$) compared with the cleanest community (annual average $PM_{2.5}$ – 6 $\mu g/m^3$) (Gauderman, Avol et al. 2004). In addition, the CHS revealed that after moving to a community with lower PM_{10} levels, an improvement was seen in the pattern of pulmonary growth (Avol, Gauderman et al. 2001). This suggests that improvements in air quality may allow children's lungs to recover from the previously experienced adverse effects (Gotschi, Heinrich et al. 2008).

CHS found significant associations between specific air pollutants (PM_{10}, $PM_{2.5}$, NO_2, O_3) and the various pulmonary function measures (FVC, FEV_1, PEF) (Peters, Avol et al. 1999). Children exposed to higher levels of PM, NO_2, acid vapor and elemental carbon, had significantly lower lung function at age 18 (Gauderman, Avol et al. 2004). Children living in communities with higher concentrations of NO_2 and PM_{10} had lungs that both developed more slowly and experienced deficits in lung function (Gauderman, McConnell et al. 2000). Days with higher ozone levels resulted in significantly higher school absences due to respiratory illness. Children with asthma who were exposed to higher concentrations of PM_{10} were much more likely to develop bronchitis (Gilliland, Li et al. 2001). Children who participated in more than 3 outdoor sports and were exposed to high levels of ozone were more likely to develop asthma than the same active children living in areas with less ozone pollution (McConnell, Berhane et al. 2002). The CHS study also showed children living within 500m of a freeway had decreases in FEV_1 compared to those living beyond 1500m of a freeway. These effects were independent of background pollutants.

There have been other studies performed within the state of California. Particulate matter (PM_{10} and $PM_{2.5}$) was associated with asthma exacerbation, measured by questionnaire and symptom diary, among African American children in Los Angeles (Ostro, Broadwin et al. 2000). In the East Bay Children's Respiratory Health Study of Traffic-related Air Pollution near busy roads, they found an increase from 5% to 8% in bronchitis symptoms and asthma symptoms in children in neighborhoods with higher concentrations of traffic pollutants (Kim 2004; Kim, Smorodinsky et al. 2004). The University of California, Berkeley Ozone Studies found significant negative associations between flow measures and O_3 in Freshman students with a low ratio of FEF_{25-75}/FVC, a marker for narrower small airways (Tager, Balmes et al. 2005).

Multi-center studies have been able to compare wide geographic areas with different pollution conditions. In the late 1980s, Schwartz examined the relationship between long-term exposure of children to air pollution and pulmonary function in the Second National Health and Nutrition Examination Survey (NHANES), which found significant decrements in pulmonary function associated with exposure (Schwartz 1989). Dockery *et al.* reported that chronic bronchitis and chest illness in children were associated with exposure to particulate air pollution in a study which compares adjusted rates across six communities in the eastern United States (Dockery, Speizer et al. 1989). Inner-City Asthma Study (ICAS), researchers examined 861 children with persistent asthma, aged 5 to 12 years, living in low-income areas in seven U.S. inner-city communities over two years: Boston, the Bronx, Chicago, Dallas, New York City, Seattle and Tucson. Results revealed that children had significantly decreased pulmonary function following exposure to higher concentrations of the air pollutants SO_2, $PM_{2.5}$, and NO_2. Higher NO_2 levels and higher levels of $PM_{2.5}$also were associated with school absences related to asthma, and higher NO_2 levels were associated with more asthma symptoms (O'Connor, Neas et al. 2008).

There have also been considerable amounts of research in other parts of the world, particularly Europe, which have demonstrated similar findings. A study in Austria found a strong association between NO_2 and asthma in 7-year olds (Studnicka, Hackl et al. 1997). In Mexico City, a highly polluted city, schoolchildren were followed for 3 years and investigators found significant yearly deficits in FVC and FEV_1 associated with 6-month means of PM_{10}, NO_2 and O_3 concentration (Rojas-Martinez, Perez-Padilla et al. 2007). A similar study in Austria reported seasonal and long-term effects of PM_{10} and O_3, however, no significant deficits in pulmonary function growth were found (Ihorst, Frischer et al. 2004). Also in Austria, a small cohort showed small improvements in pulmonary function as NO_2 levels decreased over 5 years (Neuberger and Moshammer 2004).

A recent study in Oslo, Norway modeled ambient air pollution since birth on 9- and 10-year-old children and found exposures to $PM_{2.5}$, PM_{10}, and NO_2 were associated with reduced forced expiratory flows (especially in girls), but not with forced expiratory volumes (Oftedal, Nystad et al. 2009). Two studies in Europe found significant associations between traffic density and various pulmonary function measures in school children (Wjst, Reitmeir et al. 1993; Brunekreef, Stewart et al. 2009). Some studies of SO_2, NO_2, CO and O_3 found associations with respiratory symptoms, but not pulmonary function (Hirsch, Weiland et al. 1999).

Several studies showed effects in pulmonary function associated with traffic-related exposures. In a study of children in the Netherlands, exposure to truck traffic density was associated with deficits in FEV_1, PEF and FEF_{25-75}, particularly among those who resided within 300 meters of motorways (Brunekreef, Janssen et al. 1997). In a study of children through the reunification of Germany, Sugiri observed an improvement in lung function with lower TSP and SO_2 among the children in East Germany, which then catch up the West German children 8 years after the reunification. This also points to a hope that improvements can be made during a child's development if air pollution conditions are ameliorated.

Increased respiratory tract complications in children have been associated with residence near areas of high traffic density (Brunekreef, Janssen et al. 1997; Ciccone, Forastiere

et al. 1998). Residing near busy roadways is associated with increased asthma hospitalizations, decreased lung function, and increased prevalence and severity of wheezing and allergic rhinitis (Peden 2001; Diaz-Sanchez, Proietti et al. 2003). Lin found an association between the highest tertile of traffic density and asthma, adjusted for poverty level, in a case control study of hospital admissions (Lin, Chen et al. 2002).

Holguin used road and traffic density to examine the effects of traffic on lung volumes and pulmonary function in a population with and without asthma in Mexico. For those with asthma, exhaled NO was associated with an inter-quartile increase in road density within a 50m, 100m and 200m buffer of 28%, 27% and 17%, respectively. However, the same was not true among those without asthma. Reduced FEV_1 was also associated with road density in both groups (Holguin, Flores et al. 2007).

Overall, many of these studies collected extensive covariate data including important confounders such as secondhand smoke exposure. Studies have shown that long term exposure is associated with changes in lung function in adolescents and young adults, lung function is lower in children who were live in more polluted areas and improving one's air pollution can lead to improvements in lung function. However, the statistical methods used in these studies were not designed to estimate parameters causal associations. In addition, they were difficult to synthesize and compare across various exposures, outcomes and methods.

In summary, children, especially those with asthma, are a vulnerable population at risk of suffering the adverse effects of air pollution exposure. It is unknown whether air pollution causes asthma, though short-term exposures to various air pollutants certainly exacerbate asthma. Asthma is an airway disease, characterized by airway obstruction, airway hyper-responsiveness, and mucus secretion. Air pollutants contribute to asthma burden by causing acute asthma-related symptoms and short-term declines in lung function. It is likely that a key mechanism by which air pollutants adversely impact health is through the promotion or induction of oxidative stress and inflammation. As mentioned above, O_3 and NO_2 are both reactive air pollutants that affect pulmonary function through reactions with the lining fluid of the pulmonary airways (Postlethwait, Langford et al. 1995; Pryor, Squadrito et al. 1995; Mudway and Kelly 2000). Early life events may play an important role in the development of asthma. Asthma appears to be more common in children of low birth weight and appears to be more related to growth retardation than to prematurity (Steffensen, Sorensen et al. 2000; Annesi-Maesano, Moreau et al. 2001). More will be discussed about this topic in Chapter 3.

Chapter 3: Birth Outcomes and Associations with Prenatal Air Pollution Exposure

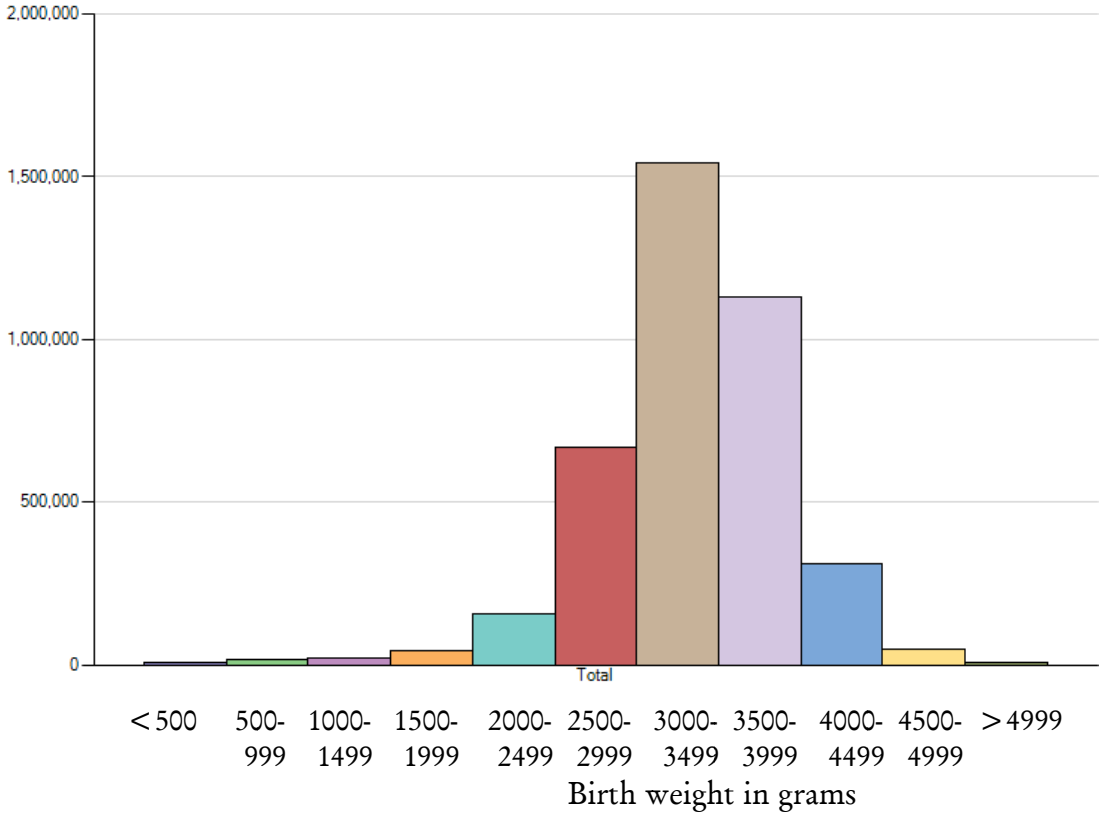
This chapter will address adverse birth outcomes such as low birth weight, preterm birth, small for gestational age, intrauterine growth retardation and the consequences of these outcomes. These adverse birth outcome, exclusive of birth defects, represent an important public health concern in terms of increased hospital costs, long-term morbidity and neonatal mortality, which will be discussed in Section 3.2. I will also discuss methodological issues related to the epidemiological study of birth outcomes and the associations of adverse birth outcomes with traffic-related air pollution.

3.1 Definitions of Adverse Birth Outcomes

3.1.1 Low Birth Weight

Low birth weight (LBW) is classified as weight less than 2500 g (5 lbs. 8 oz.). In 2006, 8.3% of infants in the United States were LBW (Martin 2009). This prevalence has increased from 6.7% in 1984 (Hamilton 2004). Although some of the increase is due to an increase in multiple births, however, even among singletons, the rate continues to grow. Low birth weight had a prevalence of 6.5% among singletons in 2006 (Martin, Brown et al. 2009). The prevalence of very low birth weight (VLBW) (less than 1500 g or less than 3 lbs. 4 oz.) among singletons was 1.1% in 2006 (Martin, Brown et al. 2009). The distribution of birth weight is essentially Gaussian with a left tail (Wilcox and Russell 1983). For example, Figure 3-1 shows the distribution of birth weight among singletons in the U.S. in 2003. African-Americans have the highest prevalence of low birth weight (11.9%) compared to other races (Martin, Brown et al. 2009). Low birth weight, although simple in its definition, is further described by gestational length as well. Term LBW and preterm LBW are often considered separate because the causes of each are distinguished between a deficit in growth versus and a shortening of gestational length.

Distribution of singleton births in the U.S. in 2003.



3.1.2 Preterm Birth

Most infants with low birth weight are born preterm (PTB), before 37 weeks (294 days) from the first day of the last menstrual period (LMP). In 2006, 11.9% of singletons were preterm (Martin, Brown et al. 2009). Preterm low birth weight is associated with perinatal mortality, neonatal mortality, childhood morbidity and nearly one half of all congenital anomalies.

3.1.3 Small for Gestational Age

Small for gestational age (SGA) is defined as birth weight and/or length at least two standard deviations below the mean for gestational age, based on data derived from a reference population (Lee, Ha et al. 2003). By definition, 2.3% of infants are SGA, however, in some cases researchers have used 3rd or 5th percentile as a cut off rather than two standard deviations (Lee, Ha et al. 2003). In contrast to preterm LBW infants, SGA infants are not small because they were born early, but rather they grew more slowly than expected for the time spent *in utero.*

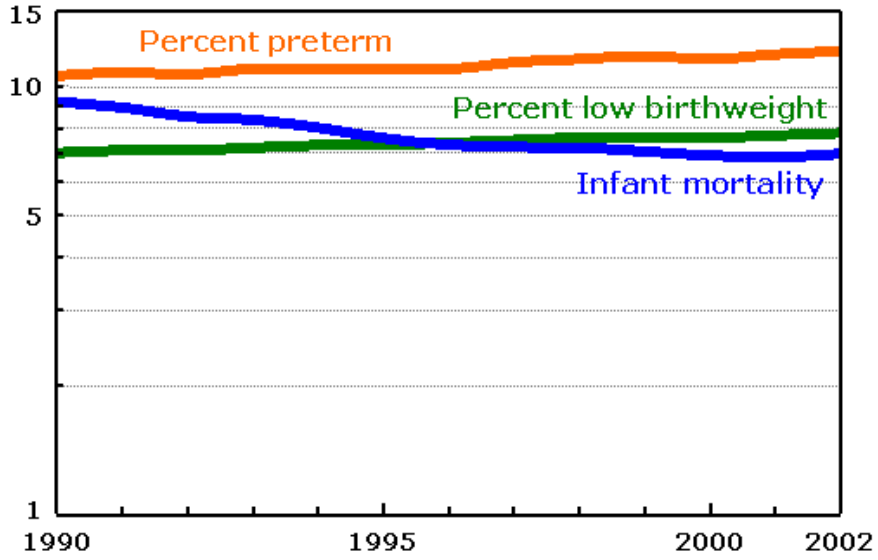
3.1.4 Intrauterine Growth Retardation

Intrauterine growth retardation (IUGR) is often defined as the 10th percentile of birth weight for gestational age. Some refer to those who are gestationally full-term ($\geq$37 weeks) but of a low birth weight ($<$2500g) as IUGR when there are not percentiles of birth weight of gestational age for a comparable population (Ritz 2008).

3.2 Burden of Adverse Birth Outcomes

The prevalence of preterm and low birth weight infants continues to increase in the U.S. (Stillerman, Mattison et al. 2008) (Behrman 2007) despite declines in LBW in the 1970s and early 1980s.

All births in the U.S. 1990-2002.



NOTE: Rates plotted on a log scale. Preterm is less than 37 completed weeks of gestation. Low birthweight is less than 2,500 grams.
SOURCE: National Vital Statistics System, NCHS, CDC.

Although recent increases in multiple births have influenced the rise, the prevalence of PTB and LBW are also continue to rise among singleton infants (Stillerman, Mattison et al. 2008). Recent studies have shown that the increased use of induction of labor and cesarean delivery has influenced the upswing in late preterm (34-36 weeks gestation) birth rate, it is not possible to know what would have happened if those births had not been induced or delivered by cesarean section (Fuchs and Wapner 2006; Bettegowda, Dias et al. 2008).

The long-term implications of adverse birth outcomes are a serious public health problem. Although ~8% of births in the U.S. are categorized as low birth weight (< 2500 grams), 65% of all infant deaths are among low birth weight infants (Wang, Ding et al. 1997). Adverse birth outcomes have a financial and emotional burden on families. The high rate of preterm births in the U.S. constitutes a public health concert that costs society at least $26 billion a year (Behrman 2007). It has been estimated that the lifelong cost per child associated with being a low birth weight infant is $436,000 (Wong, Gohlke et al. 2004). Although an increase in adverse birth outcomes is viewed as a public health problem, it is likely that the increase in preterm and low birth weight births is, in part, a result of improved neonatal care. Births that in the past would have resulted in infant mortality may now be saved as preterm or low birth weight infants.

3.3 Fetal Origins of Asthma

Low birth weight and preterm birth are associated with increased rates of coronary heart disease and the related disorders, stroke, hypertension and type 2 diabetes in adulthood. The associations, in support of the fetal origins hypothesis, are thought to be consequences of developmental plasticity, meaning environmental conditions during development can affect physiological and morphological states by programming growth (Barker 2004). It has been suggested that the fetal period and early childhood are both critical times for the development of asthma and other allergic diseases (Bjorksten 1999). It is possible the growth of the lung is altered due to these fetal exposures.

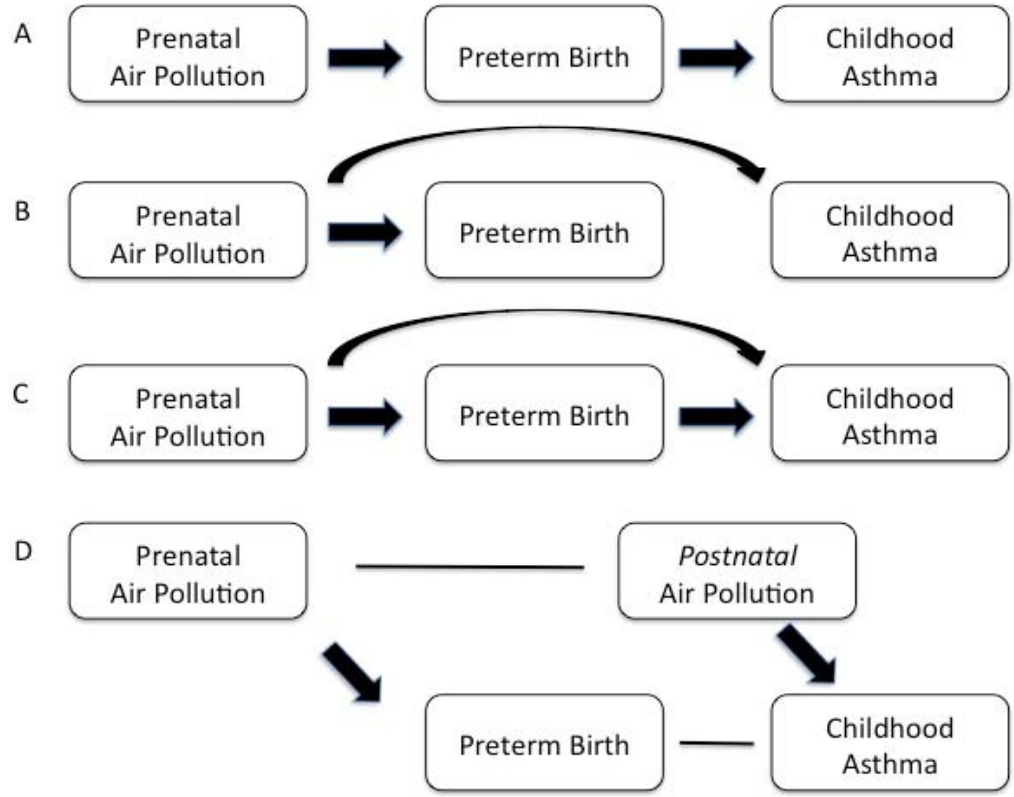
In the fetal origins paradigm, birth weight is a marker for many developmental processes and the true etiologic pathways are unknown (Gillman 2002). To be causal, exposure must precede the outcome. For preterm birth, we do not know when the mechanism for preterm birth begins, other than that it predates the delivery date by some amount of time. Therefore, it is difficult to determine if there is a critical period of exposure during pregnancy, which may cause PTB or LBW. Reduced duration of pregnancy can be considered an indicator of disturbance in fetal development. Although the causes of preterm birth are diverse, these disturbances in fetal development predispose individuals to diseases later in life. However, it is difficult to disentangle the pathway from supposed genetic and environmental factors to later diseases.

Many studies have examined the relationship between preterm birth and development of asthma. Most have concluded that preterm infants have a small but significant increased risk of asthma compared with term infants (Jaakkola, Ahmed et al. 2006). The majority of studies have been cohort studies, though some have been cross-sectional. In most studies, asthma was confirmed by physician diagnosis, yet some examined the relationship with asthma symptoms, which may be more specific, but less reliable.

Schwartz found an association between LBW and recurrent wheeze in children in the U.S. (Schwartz, Gold et al. 1990). In a cohort in Denmark, one study observed no association between LBW nor PTB and asthma (Steffensen, Sorensen et al. 2000). Jaakkola found an increase in odds of asthma at age 7 years among those born LBW and PTB (Jaakkola and Gissler 2007). In a recent study, Latzin *et al.* found that prenatal exposure to NO_2 and PM_{10} was associated with higher respiratory need and airway inflammation in newborns as measured by minute ventilation (Latzin, Roosli et al. 2009).

A few models for potential causal pathways were proposed in Jaakkola's systematic review. Figure 3-3 shows a directed acyclic graph (DAG) of potential explanations and pathways of causation. One hypothesis is that preterm delivery is on the causal pathway between environmental factors such as prenatal air pollution, in this report, and asthma (A). A second proposal is that prenatal air pollution causes both preterm delivery and asthma (B). It is also possible that both scenarios are true (C). Finally, it is possible that prenatal air pollution causes preterm delivery and postnatal air pollution causes asthma, forcing an association between prenatal air pollution exposure and asthma if prenatal and postnatal exposures are strongly correlated (D).

Directed acyclic graph of potential explanations and pathways of causation



3.4 Methodological Issues with the Study of Birth Outcomes

Adverse birth outcomes result from diverse etiologic pathways. Established risk factors for low birth weight include maternal smoking, alcohol use, lack of prenatal care, infection, malnutrition, placental factors, and fetal factors (*e.g.*, chromosomal abnormalities, genetic defects, growth hormone deficiency or short stature syndromes) (Lee, Ha et al. 2003). Additional factors are associated with adverse birth outcomes including maternal age (<18, >35), height and weight, race and ethnicity (particularly African-Americans), single marital status, low socioeconomic status, parity, previous LBW or PTB, low weight gain during pregnancy, hypertension and diabetes. Some of these factors such as low socioeconomic status and race are potential confounders and/or effect modifiers in the relationship between air pollution exposure and adverse birth outcomes.

The quality of birth certificate data vary and certain fields are incomplete or prone to measurement error. The record of gestational age on the birth certificate is often based on the maternal recall of her last menstrual period. In some cases, it may be corrected or altered by ultrasound measurements. However, this measurement is less precise than that of birth weight, for example. Socioeconomic status is an example of one factor that is not recorded.

Some studies use maternal education or whether the birth costs were paid by Medicaid (Medi-Cal) as proxies for socioeconomic status.

There are well-documented differences in the frequency of adverse birth outcomes across racial groups, which are not driven solely by variations in socioeconomic status. Maternal genetic polymorphisms that may modify the risk of these adverse birth outcomes have been identified (Engel, Erichsen et al. 2005; Engel, Olshan et al. 2005). Moreover, environmental exposures such as second hand tobacco smoke and benzene may interact with these genetic polymorphisms (including GST mentioned above) to further increase the risk for low birth weight (Wang, Zuckerman et al. 2002) and preterm birth(Nukui, Day et al. 2004). In addition, polymorphisms in genes have been associated significantly with increased risk of preterm birth (Annels, Hart et al. 2004), (Adams and Eschenbach 2004). Studies such as these highlight the importance of identification of gene-environment interactions. Although these factors are currently difficult to control for in epidemiological studies, as technology and knowledge of genetics improve, future studies should aim to incorporate these risk factors and potential effect modifiers.

3.5 Potential Biologically Plausible Mechanisms

There are a number of potential mechanisms through which ambient air pollution may affect LBW and PTB. Air pollution may affect maternal respiratory or general health and, in turn, impair uteroplacental and umbilical blood flow, transplacental glucose and oxygen transport, all known as major determinants of fetal growth (Ritz and Wilhelm 2008). Additionally, the pollutants to which the fetus is exposed may cause oxidative stress (which can affect the embryo in the earliest phase of growth), inflammation of pulmonary and placental cells (which can induce DNA damage), and changes in blood coagulation or hemodynamic responses (Kannan, Misra et al. 2006; Kannan, Misra et al. 2007; Ritz and Wilhelm 2008).

Although maternal smoking is not a central point in this paper, it should be mentioned for several reasons. First, it is the leading environmental risk factor for an adverse pregnancy outcome and is a well-studied example of a prenatal exposure that has a profound effect on infants. Infants born to mothers who smoke are on average ~200 grams smaller, on average, than the infants of non-smokers. Smokers are at high risk of delivering very small preterm infants and their infants have higher perinatal mortality at every relative birth weight (Wilcox 1993). Carbon monoxide, a major constituent of tobacco smoke, crosses the placenta rapidly and is detected in the fetal circulation at levels 15% higher than maternal levels (Longo and Ching 1977; Ritz and Wilhelm 2008).

Moreover, second-hand smoke (SHS), which elevates indoor levels of CO and PM_{10}, is associated with reduced birth weight and small for gestational age (Ritz and Yu 1999). These associations are not as consistent as those well-established associations with maternal smoking; however, studies are suggestive of a relationship. One study found passive exposure of pregnant women to SHS during the third trimester to be positively associated with asthma- and allergy-related symptoms in their preschool age children (Xepapadaki, Manios et al. 2009).

Additionally, it is hypothesized that the potential mechanism of action for the effects of maternal smoking are the same as those for maternal exposure to air pollution. Suggested

mechanisms include disturbances of the pituitary-adreno-cortico-placental system and uterine blood flow, and/or increased maternal susceptibility to infections. Through epigenetic pathways described elsewhere (Liu, Ballaney et al. 2008; Baccarelli, Wright et al. 2009; Perera, Tang et al. 2009) air pollution may trigger premature contractions or premature rupture of the membranes leading to preterm birth. Inhalation and absorption of compounds contained in ambient air mixtures, such as polycyclic aromatic hydrocarbons (PAHs) from vehicular exhaust, could interfere with the development and nutrition of the fetus and cause fetal distress, and recent data suggest that they may lead to epigenetic modification of fetal DNA (Perera, Tang et al. 2009). Increases in particulate matter concentrations have been shown to increase blood pressure and heart rate in healthy adults; thus, particles may also be able to affect uterine blood flow (Brook 2007). Ultrafine particles may escape phagocytosis by alveolar macrophages and translocate to extrapulmonary organs (Oberdorster and Utell 2002) and transfer potentially toxic compounds– such as (PAHs), organic acids, transition metals – to the fetus and the placenta. These compounds may interfere with placental development and subsequent nutrient and oxygen delivery to the fetus (Dejmek, Solansky et al. 2000). In addition to direct effects on DNA, many of these compounds lead to the production of reactive oxygen and nitrogen species that can cause further DNA damage and provoke inflammation.

The well-accepted link between maternal smoking and adverse birth outcomes lends support for a role of ambient air pollution impacts on pregnancy (Ritz and Wilhelm 2008). Moreover, there may be an effect modification between maternal smoking in pregnancy and exposure to air pollution.

3.6 Previous Studies on the Effect of Prenatal Air Pollution Exposure on Birth Outcomes

In the past decade, a large number of studies have examined the associations between of traffic-related air pollution exposures on birth outcomes (low birth weight, preterm birth, birth defects and infant mortality). There have been more studies for low birth weight than for preterm birth, though the results are similar. It continues to be uncertain when the most vulnerable time in pregnancy is for the influence of air pollution. Similarly, associations have been found with all of the criteria pollutants but little consistency as to which ones are the most critical. This variability in results may be due to the variability in the locations where these pollutants are measured. Although most pollutants are evaluated individually, the mixtures of many pollutants vary across the country and across the world.

Ritz, *et al.* examined the associations between ambient air pollution and outcomes among births in Los Angeles, CA from 1989-1993 (Ritz 1999, 2000). Birth certificates were used in a retrospective cohort study of LBW and PTB. Exposures of air pollutants, including CO, NO_2 and PM_{10}, were measured by monitoring stations closest to the mothers' residence. For the birth weight analysis, Ritz *et al.* excluded multiple births and those shorter than 37 or longer than 44 weeks gestation and birth weights greater than 5500 grams and less than 1000 grams. Women with hypertension, diabetes, and uterine bleeding during pregnancy were also excluded from the analysis. Using logistic regression, they found that those exposed to high levels of CO ($>95^{th}$ percentile; 5.5 ppm) during the third trimester had a relative risk of 1.22

(95% CI: 1.03-1.44) for LBW compared with those exposed to low levels of CO ($<50^{th}$ percentile) (Ritz and Yu 1999).

In the same study, they also found associations between preterm birth and high PM_{10} during the first month (RR = 1.16, 95% CI: 1.06-1.26) and in the last six weeks before birth (RR = 1.20, CI:1.09-1.33). In this analysis, Ritz *et al.* included births between 26 and 44 weeks gestation (Ritz, Yu et al. 2000). Additionally, in the southern California cohort, Wilhelm also showed an increase in preterm birth for those exposed to higher amounts of traffic in winter months, with higher associations among those with low socioeconomic status (Wilhelm and Ritz 2003). Traffic was defined as categorical values of distance weighted traffic density (DWTD),which consist of annual average daily traffic counts ($<20^{th}$ percentile, 20-80^{th} percentile, and $>80^{th}$ percentile). Socioeconomic status was assessed by unemployment, family poverty and income from public assistance obtained from the 1990 Census at the county level (Wilhelm and Ritz 2003).

In a systematic review of the effects of air pollution on pregnancy outcomes, Ghosh *et al.* evaluated infant sex as a potential effect modifier. Across a range of studies of various pollutants and metrics, they found that females had a slightly higher risk of low birth weight (OR's ranged from 1.07-1.62) and males had a slightly higher risk of preterm birth (OR's ranged from 1.11-1.20) (Ghosh, Rankin et al. 2007). This dissertation will not address sex as an effect modifier, though future analyses of these data will include exploration of effect modification by sex, race and socioeconomic status.

The analyses in this dissertation will not address infant mortality nor birth defects, but these birth outcomes will be described due to common biological mechanisms for the harmful effects of traffic-related air pollution. Several studies have found a relationship between traffic-related pollution on infant mortality (Woodruff, Grillo et al. 1997; Pereira, Loomis et al. 1998; Bobak and Leon 1999; Lipfert, Zhang et al. 2000; Ritz, Wilhelm et al. 2006; Woodruff, Parker et al. 2006; Loomis, Castillejos et al. 1999). The strongest association was found during the postneonatal period, however, it is difficult to disentangle the effects during pregnancy and those during the postneonatal period, particularly since the mothers' exposure during pregnancy is often very similar to that of the infant's exposure after birth. Ritz *et al.* found a higher odds of birth defects associated with higher exposure to CO during the second month of pregnancy (Ritz, Yu et al. 2002). Strickland also found an increased risk of a specific birth defect, patent ductus arteriosus, associated with increases in PM_{10} (Strickland, Klein et al. 2009). As mentioned by Ritz (Ritz and Wilhelm 2008), although the effect sizes for pollutants are quantitatively lower than those in studies of maternal smoking, exposure to ambient air pollution occurs broadly across large populations. Furthermore, it is far less modifiable by the individual, and thus of great public health and policy importance.

Although this paper focuses on studies in California, investigations have been done in the Czech Republic, South Korea, Australia, Lithuania, Germany, China, Canada, and other states in the U.S. including Georgia, Massachusetts, Connecticut, Pennsylvania, Maryland and Washington D.C. (Xu, Ding et al. 1995; Bobak and Leon 1999; Dejmek, Selevan et al. 1999; Bobak 2000; Rogers, Thompson et al. 2000; Maisonet, Bush et al. 2001; Maroziene and Grazuleviciene 2002; Ha, Lee et al. 2003; Lee, Ha et al. 2003; Liu, Krewski et al. 2003; Woodruff, Parker et al. 2003; Yang, Chang et al. 2003; Sagiv, Mendola et al. 2005; Dugandzic,

Dodds et al. 2006; Leem, Kaplan et al. 2006; Bell, Ebisu et al. 2007; Jalaludin, Mannes et al. 2007; Slama, Morgenstern et al. 2007; Brauer, Lencar et al. 2008; Genereux, Auger et al. 2008; Parker, Mendola et al. 2008; Zeka, Melly et al. 2008; Jedrychowski, Perera et al. 2009; Suh, Kim et al. 2009; Wang, Ding et al. 1997) Further studies of the effects of air pollution on low birth weight and preterm birth are summarized in Table 3-4, below.

Prenatal traffic-related air pollution exposure and low birth weight and preterm birth outcomes.

Reference	Study Population	Exposure	Outcome	Results	Additional information
Xu, X. (1995)	Beijing, China 1988 N=25,370	SO_2 TSP	PTB	SO_2 OR=1.21 TSP OR=1.10	per 100 μg/m_3 inc
Wang, X. (1997)	Beijing, China 1988-1991 N=74,671	SO_2 TSP	LBW	SO_2 OR=1.11 TSP OR=1.10 (per 100 μmg/m_3 inc)	3rd tri; adj for gestational age, residence, year of birth, maternal age, infant sex
Dejmek, J. (1999)	Czech Republic 1994-1996 N=1,753	PM_{10} (1st mo) $PM_{2.5}$ (1st mo)	IUGR	$PM_{10}OR_{med}$=1.62 $PM_{10}OR_{high}$=2.64 $PM_{2.5}OR_{med}$=1.26 $PM_{2.5}OR_{high}$=2.11	Low, ref (1st tertile) Medium (2nd tertile) High (3rd tertile)
Ritz, B. (1999)	Los Angeles, CA 1989-1993 N=125,573	CO	LBW	OR=1.22 (3rd tri)	>95th percentile (5.5 ppm 3-mo avg); adj commuting habits, infant sex, prenatal care, maternal age, ethnicity education
Bobak, M. (1999)	Czech Republic 1986-1988	O_2 TSP NO_x	LBW S	SO_2 OR_{LBW}=1.10 TSP OR_{LBW}=1.04 (NS) NO_x OR_{LBW}=1.07 (NS)	Ecologic study in 45 districts; 50 μg/m^3 increase in pollutant; adj for SES factors
Bobak, M. (2000)	Czech Republic 1990-1991 N=108,173	SO_2 (1st tri) TSP (1st tri) NO_x(1st tri)	LBW PTB IUGR	SO_2 OR_{LBW}=1.20 SO_2 OR_{PTB}=1.27 TSP OR_{LBW}=1.15	50 μg/m^3 increase in pollutant; adj for SES factors

Ritz, B. (2000)	Los Angeles, CA 1989-1993 N=97,518	CO NO_2 PM_{10}	PTB	PM_{10} RR=1.20 (6wks before birth) PM_{10} RR=1.16 (1^{st} mo) CO RR=1.13 (6wks before birth; inland)	50 μg/m^3 increase in PM_{10}; 3 ppm increase in CO; adj for maternal age, education, parity, smoking, interval since live birth, prenatal care, infant sex, season of birth or conception, previous LBW, PTB
Rogers, J. (2000)	Georgia 1986-1988	SO_2 TSP	VLBW (<1500 g)	OR=2.88 >95^{th} %-ile (56.75μg/m^3)	Adj for race, toxemia, ETS, maternal weight gain, prepreg weight, age, prenatal care, income, education, drug use, alcohol, stress, infant sex, paternal education
Ha, E. (2001)	Seoul, South Korea 1996-1997 N = 3,310	CO (1^{st} tri) NO_2 (1^{st} tri) SO_2 (1^{st} tri) TSP (1^{st} tri)	LBW	CO RR=1.08 NO_2 RR=1.07 SO_2 RR=1.06 TSP RR=1.04	IQ increase in pollutant; adj for gestational age, maternal age, parental education level, parity, infant sex
Maisonet, M. (2001)	Massacusetts, Connecticut, Pennsylvania, Washington, D.C. 1994-1996 N=89,557	CO PM SO_2	LBW	CO OR=1.31 (3^{rd} tri) SO_2OR_{4thQ}=1.21 (2^{nd} tri) SO_2OR_{3rdQ}=1.20 (2^{nd} tri) SO_2OR_{2ndQ}=1.21 (2^{nd} tri) SO_2(ref 1^{st} Q)	Adj for maternal age, race, smoking, alcohol, first born, infant sex, marital status, previous termination, prenatal care, weight gain, gestational age, season of birth
Maroziene, L. (2002)	Kaunas, Lithuania 1998 N=3988	NO_2	LBW PTB	OR_{LBW}=1.25 OR_{PTB}=1.14 (2^{nd}tertile) OR_{PTB} =1.68 (3^{rd}tertile)	10 mmg/m^3 increase in NO_2

Lee, B. (2003)	Seoul, South Korea 1996-1998 N = 13,835	CO PM_{10} SO_2 NO_2	LBW	Small and borderline significant results for 1st tri, 2nd tri and entire preg	IQ increase in pollutant; evaluated each month, each tri and entire preg
Liu, S. (2003)	Vancouver, Canada 1985-1998 N = 229,085	SO_2 NO_2 CO	LBW PTB IUGR	SO_2 OR_{LBW} = 1.11 (1st mo) SO_2 OR_{PTB} = 1.09 (last mo) CO OR_{PTB} = 1.08 (last mo) SO_2 OR_{IUGR} = 1.07 (1st mo) NO_2 OR_{IUGR} = 1.05(1st mo) CO OR_{IUGR} = 1.06(1st mo)	Adj maternal age, parity, infant sex, gestational age, BW, mo of birth
Wilhelm, M. (2003)	Los Angeles, CA 1994-1996 N = 65,379	DWTD (traffic)	LBW (term/ preterm) PTB	OR_{tLBW} = 1.39 OR_{ptLBW} = 1.24 RR_{PTB} = 1.15 (3rd tri - fall/winter) RR_{PTB} = 1.08	Highest quintile of DWTD; adj for maternal age, race/ethnicity, education, parity, interval since previous live birth, prenatal care, infant sex, previous LBW or PTB, season of birth, year of birth
Woodruff, T. (2003)	263 counties in U.S. 1998-1999 N = 7,355,696	CO PM_{10} SO_2 NO_2 O_3	SGA PTB	OR_{PTB} = 1.04 (per unit increase	County level air pollution; adj for region, age, parity, marital status, education, race/ethnicity
Mohorovic, L. (2004)	1987-1989 Croatia	SO_2	PTB LBW	Shorter gestation -0.09 (1st mo) -0.08 (2nd mo) Lower birth weight -0.08 (1st mo) -0.07 (2nd mo)	

Ponce, N. (2005)	Los Angeles, CA 1994-96 N = 37,347	DWTD (traffic)	PTB	Traffic density > 80th %-iles 20th%-ile associated with PTB more among low SES in winter months; OR = 1.30	Adj for poverty, unemployment, and dependence on income from public assistance (census tract)
Sagiv, S. (2005)	Pennsylvania 1997-2001 N = 187,997	PM_{10} SO_2	PTB	PM_{10} RR = 1.07 (NS) per 50 μg/m^3 SO_2RR = 1.15 (NS) per 15 ppb inc	Time-series analysis; exposure 6 wks before birth
Wilhelm, M. (2005)	Los Angeles, CA 1994-2000 N = 106,483	CO PM_{10} $PM_{2.5}$	termLBW PTB	CO OR_{LBW} = 1.36 (3rd tri) CO OR_{PTB} = 1.18 (1st tri) (w/in 1 mile of monitoring stations)	Highest vs lowest quartile of exposure; adj for maternal age, race, education, parity, interval since previous live birth, prenatal care, infant sex, previous LBW or PTB; adj for gestational age (LBW)
Salam, M. (2005)	California 1975-1987 N = 3901	O_3 CO NO_2 PM_{10}	BWT IUGR	O_3 (12 ppb inc) -47.2 g CO (1.4 ppm inc) -21.7 g O_3 OR_{IUGR} = 1.2	PM_{10} 3rd tri -21.7 g O_3 3rd tri OR_{LBW} = 1.4 NO_2 1st tri OR_{IUGR} = 1.2 CO 1st tri OR_{IUGR} = 1.2
Dugandzic, R. (2006)	Nova Scotia, Canada 1988-2000 N = 74,284	PM_{10} (1st tri) SO_2 (1st tri) O_3	LBW	PM_{10} RR_{crude} = 1.33 SO_2RR_{crude} = 1.36	Results (adj for YOB) were attenuated (NS); dose response noted for SO_2; NS for O_3

Leem, J-H. (2006)	Incheon, Korea 2001-2002 N=52,113	CO PM_{10} SO_2 NO_2	PTB	CO $RR_{1st\ tri}=1.26$ PM_{10} $RR_{1st\ tri}=1.27$ SO_2 $RR_{1st\ tri}=1.21$ NO_2 $RR_{1st\ tri}=1.24$ CO $RR_{3rd\ tri}=1.16$ NO_2 $RR_{3rd\ tri}=1.21$	Highest vs lowest quartile of exposure; adj maternal age parity, sex, season of birth, parental education level
Rogers, J. (2006)	Georgia 1986-1988 N=325	PM_{10}	VLBW PTB	$OR_{VLBW}=2.54$ $OR_{PTB}=4.31$	Case control; compared to term/AGA controls; adj for maternal age, race, education, smoking, prepreg weight, weight gain, toxemia, anemia, asthma, birth season
Huynh, M. (2006)	California 1999-2000 N=42,692	$PM_{2.5}$ CO	PTB	$PM_{2.5}$ OR=1.15 (entire preg) OR=1.21 (1^{st} mo) OR=1.17 (last 2 wks)	Matched case control study; adj for maternal race/ethnicity, parity, marital status, education; results for CO - NS
Bell, M. (2007)	Massachusetts, Connecticut 1999-2002 N=358,504	PM_{10} $PM_{2.5}$ NO_2 CO SO_2	BWT	PM_{10} -8.2g $PM_{2.5}$ -14.7g NO_2 -8.9g CO -16.2g SO_2 -0.9g (NS)	change in BW for IQ increase in pollutant; higher among blacks; adj for gestational age, prenatal care, type of delivery, child's sex, birth order, weather, year, mother's race, education, marital status, age, tobacco use; lower bwt was associated with 3^{rd} tri PM_{10}, 1^{st} and 3^{rd} tri CO, 1^{st} tri NO_2 and SO_2, 2^{nd} and 3^{rd} tri $PM_{2.5}$

Jalaludin, B. (2007)	Sydney, Australia 1998-2000 N=123,840	O_3 SO_2 NO_2	PTB	O_3 (1st tri) and SO_2 associated with PTB in some analyses	Inconsistent results; association with spring conception
Slama, R. (2007)	Munich, Germany 1998-1999 N=1,016	$PM_{2.5}$ $PM_{2.5}$ absorbance NO_2 LUR (traffic)	BWT < 3kg	$PM_{2.5}$ PR=1.7 $PM_{2.5}$ absorbance PR=1.8 NO_2 PR=1.2 (NS)	Highest vs lowest quartile of exposure
Brauer, M. (2008)	Vancouver, Canada 1999–2002 N=70,249	CO NO_2 O_3 SO_2 PM_{10} $PM_{2.5}$ BO NO Residing w/in 50m of highway	SGA LBW PTB	w/in 50m highway OR_{SGA}=1.26 OR_{LBW}=1.11(NS)	Population-based cohort study; low ambient air pollution profile; all pollutants (except O_3) showed small but consistent association with SGA
Généreux, M. (2008)	Montréal, Canada 1997- 2001 N = 99,819	Residing w/in 200m of highway and neighborhood SES (census tract)	PTB	High SES and residing w/in 200m of highway OR_{PTB}=1.58 OR_{LBW}=1.81 OR_{SGA}=1.32	Adjusted for individual maternal education, maternal active and passive smoking, nutritional status, domestic violence and access to prenatal care
Yang, CY. (2003)	Taiwan, China 1992-1997 N = 6,251	Distance from freeway		OR=1.30 (w/in 500m freeway)	w/in 500m from freeway compared to 500-1,500 m

Zeka, A. (2008)	Massachusetts 1996-2002 N=425,751	CADT (traffic) Distance to roadways	LBW SGA PTB	CADT $OR_{SGA}=1.02$ CADT -2 g (bwt) (reduced precision with SES modifiers)	Adj for median household income (census tract) and maternal education, SEM, ethnicity, smoking
Parker, J. (2008)	Utah Valley 1986-1987 N=39,051	PM_{10}	PTB	Slight reduction in PTB during steel mill closure (NS), stronger reduction during 2nd tri	Steel mill closure used as natural experiment
Jedrychowski, W. (2009)	N=481	$PM_{2.5}$	Fetal growth	IR change in pollutant associated with a change in bwt -97.2 g (NS) length -0.7 cm male bwt -189 g female bwt -17g male length -1.1cm	Adj for maternal education, height, prepreg weight, ETS, parity, season of birth, gestational age, sex of infant
Suh, Y. (2009)	Seoul, South Korea 1998-2000 N=374,167	PM_{10}	PTB	OR=1.09 (3rd tri)	Survival analysis; IQ increase in PM_{10}

ADJ = adjusted
BC = black carbon
BWT = birth weight
CADT – cumulative average daily traffic
CO = carbon monoxide
DWTD = distance weighted traffic density
ETS = environmental tobacco smoke
IQ = interquartile
LBW = low birth weight
LUR = land use regression
MO = month
NO = nitrous oxide
NO_2 = nitrogen dioxide
NO_x = nitrogen oxides
NS = not significant
OR = odds ratio
$PM_{2.5}$ = particle matter with aerodynamic diameter <2.5 μm
PM_{10} = particle matter with aerodynamic diameter <10 μm
PREG = pregnancy
PTB = preterm birth
RR = relative risk
SES = socioeconomic status
SGA = small for gestational age
SO_2 = sulfur dioxide
TSP = total suspended particles
VLBW = very low birth weight
WKS = weeks
YOB = year of birth

3.7 Limitations of the Current Literature on the Study of Air Pollution and Birth Outcomes

The epidemiology of air pollution and birth outcomes continues to evolve. If a relationship exists, it is subtle and not easy to characterize due to many methodological challenges (Stillerman, Mattison et al. 2008). Despite a growing body of evidence for adverse health effects of traffic-related pollutants, further study is warranted to address some of the current limitations described below.

Studies have aimed to identify specific periods during gestation that may be critical to growth. As presented in the table above, some studies have divided the pregnancy into trimester, others into months, and some have looked at the 2 or 6 weeks prior to birth. For LBW and PTB, the first and third trimesters have been implicated as the most relevant (Ritz and Wilhelm 2008). For example, fetal growth is more variable during the third trimester, which may be hypothesized to have the greatest association. However, growth can also be affected through placentation, which occurs in the first trimester. The trimester exposure is often dependent on the season in which it takes place, but there is also colinearity between a given trimester and the entire pregnancy, which makes it difficult to sort out the critical period if there is one. Additionally, it must be considered that despite a potential critical period, cumulative exposure across the entire pregnancy may affect birth outcomes.

While air monitors placed in urban areas allow estimation of community-wide average exposures, they are less apt to provide quality data for spatially heterogeneous air pollutants such as CO and $PM_{2.5}$ where proximity to sources may result in very high exposures (Ritz and Wilhelm 2008). Carbon monoxide is a spatially heterogeneous pollutant and levels measured at monitoring stations may only reflect concentrations within a small distance to the monitor. Particulate matter pollution is comprised of a mixture of pollutants. This mixture also varies from location to location. This variability in composition could explain the inconsistency of results across different studies. As mentioned above, pollutants are highly correlated, which makes it difficult to estimate the sole effect of any individual pollutant.

Pollution monitors are not sited everywhere people live and do not always provide continuously measured data. For example, $PM_{2.5}$ is measured every 3-6 days while CO and O_3 are measured hourly. Exposure assessment models use the spatial and temporal data to predict air pollution in places and at times where it is not measured; however, dependence on these models may lead to misclassification that is not uniform across space or time. This differential misclassification could bias the effect estimate in an unknown direction. Zeger *et. al.*, examined the effect of air pollution exposure measurement error on estimates of mortality in time-series studies. They found that classical errors, that is, the difference between measured ambient levels of pollution and the true values of the individuals to which the measurements are assigned, are the likely source of substantial bias. These errors can be quantified in studies designed to measure them and the errors can be incorporated into models in analysis (Zeger, Thomas et al. 2000).

The majority of studies examine the influence individual pollutants, yet it is the complex mixture of pollutants to which the mother and fetus are exposed. Few studies have addressed multi-pollutant models due to high colinearity among pollutants (Bell 2007). Source apportionment, such as receptor models and air-quality dispersion models, allows evaluation of the effects of several sources that contribute to a given ambient environment (HEI 2010).

Another strategy is to study the source itself, such as traffic counts, which addresses the mixtures of multiple pollutants.

The studies that have been done are located in a diverse group of geographical locations making comparisons difficult. The pollution profiles vary in different locations and factors affecting air pollution and health complicate the synthesis of these studies. Season of exposure can play an important role in the development of exposure estimates to ambient pollutants. Season can have an effect on temperature, weather, allergy, food availability, environmental exposure (such as pesticides or water quality).

In studies where the mother's residence is used as the point of reference for environmental exposures during pregnancy, this location may be problematic. It is assumed that the residence at birth is the same for that of the entire pregnancy and that the mother spent a considerable amount of time at the residence. It is possible this assumption may be incorrect, particularly for early trimesters. Time activity patterns and work locations may be able to reduce this misclassification.

There are factors that may confound and/or modify the relationship between traffic-related air pollution and birth outcomes such as socioeconomic status (SES), race and maternal smoking. Some studies have even found evidence that SES is an effect modifier (Wilhelm and Ritz 2003; Ponce, Hoggatt et al. 2005). Wilhelm *et al.* found an association between traffic-related air pollution and preterm birth in low and middle-income neighborhoods, but not in the higher-income neighborhoods (Wilhelm and Ritz 2003). However, this study did not include maternal smoking as a potential confounder and maternal smoking is a known cause of adverse birth outcomes and is likely to be associated with low socioeconomic status.

Most studies of air pollution during pregnancy and birth outcomes only adjust for risk factors on the birth certificate. Some studies have also compiled additional data from the census at the census tract level (Ponce, Hoggatt et al. 2005; Wilhelm and Ritz 2005),the zip code level (Ritz and Yu 1999) or block group level (Wilhelm and Ritz 2003). One study obtained additional information from an interview/questionnaire on a subsample of the study population (Ritz, Wilhelm et al. 2007).

In summary, low birth weight and preterm birth are important markers of health both in the neonatal period, through childhood and possibly into adulthood. Traffic-related pollution may play a role the complex etiology of low birth weight and preterm birth. It is still unclear which specific pollutants or mixtures or specific sources, at what levels and during which time periods during pregnancy, are the most critical for overall growth and lung development.

3.8 Statistical Approaches in the Study of Air Pollution and Birth Outcomes

Statistical methods applied to answer these questions have in large part included logistic regression estimating the odds of an adverse outcome conditional on the prenatal exposure to air pollution and on covariates including maternal age, ethnicity, socioeconomic status, and in some cases smoking status. These reported odds ratios must be interpreted as stratum-specific estimates and cannot be considered as marginal estimates of population estimates on pregnancy women.

The commonly used statistical methods assume the models are properly specified and do not explicitly discuss the potential for extrapolation beyond the scope of their data. This

will be discussed in further detail in the following chapter (4). Finally, there have been no studies on the associations between ambient air pollution and birth outcomes that have specified a causal model and used causal inference methods.

In the next chapter, I will discuss the limitations of these statistical methods and present alternative statistical methods that will address some of these limitations and highlight challenges of all statistical methods. In Chapters 6, I will revisit birth outcomes in an analysis of the SAGE study. The analysis uses causal inference methods with data from the birth certificate registry from years 2000-2006, traffic density measurements across four counties in the Central Valley of California, and 2000 U.S. census variables measured at the block group level for these areas. This large sample size across a diverse geographic range of California with highly refined measurements of traffic exposure will provide an estimate of the causal associations between traffic exposure during pregnancy and the outcomes of low birth weight and preterm birth in this population.

Chapter 4: Methods

4.1 Study Populations

This dissertation uses data from two different studies to estimate causal associations between prenatal air pollution and children's health outcomes.

4.1.1 Fresno Asthmatic Children's Environment Study – Lifetime Exposure

The Fresno Asthmatic Children's Environment Study (FACES) is focused on the evaluation of the association between responses to short-term increases in air pollutants and bioaerosoles and asthma severity and growth of lung function in a cohort of children with asthma in Fresno, California. This community is notable for a high prevalence of asthma among an ethnically diverse population, and for high levels of ambient air pollution, especially PM. Air pollution-related asthma morbidity is a major public health problem in the study area and in the state of California, and the nation as a whole. The San Joaquin Valley remains a non-attainment area for O_3, $PM_{2.5}$ and PM_{10} (CARB 2010)(Appendix 3). Children with asthma were enrolled at age 6-11 years. Spirometry was performed approximately every 6 months, providing a longitudinal record of lung function as the children aged. This study is discussed in more detail in Chapter 5.

As part of this research, an ancillary study was added to extend the ambient air pollution exposure period to the time of pregnancy and early life. This study, obtained estimates based on residential interpolation from existing monitors of ambient air pollution for each trimester, the entire pregnancy and first six years of life for the residences that the mother and child lived during these times. Early in the FACES study, the parents self-reported the street addresses for all of these locations. This study allowed the assessment of the influence of different exposure time periods that individually and cumulatively affect pulmonary function in childhood. In Chapter 5, I will discuss how data from the FACES cohort and the FACES-LITE sub-study were used to evaluate the causal associations between exposure to ambient air pollution during pregnancy and the pulmonary function at ages 6 to 13 years.

4.1.2 Study of Air pollution, Genetics and the Early life events

The Study of Air pollution, Genetics and Early life events (SAGE) was designed to investigate the causal associations between ambient air pollution during pregnancy and birth outcomes. Highly refined exposure data from California Regional PM_{10}/ $PM_{2.5}$ Air Quality Study (CRPAQS) (Hyslop, Brown et al. 2003) was linked to California birth certificate data, provided by the California Department of Health Services (Sacramento, CA). This study encompassed an even larger geographic area of four counties in the San Joaquin Valley of California (Kern, Fresno, Stanislaus, San Joaquin) to evaluate the impact of exposure to ambient air pollution during pregnancy on adverse birth outcomes. The SAGE population sample incorporated births from years 2000-2006. Additional data were compiled from the 2000 census to provide contextual information on socioeconomic status (SES) at the block group level. In Chapter 6, I will discuss how I examined the causal association between residential traffic density as measured by California Department of Transportation and low birth weight, defined as <2500g, reported on the birth certificate.

4.2 Statistical Methods

Randomized controlled trials (RCTs) are often considered the "gold standard" for study designs to provide causal inference with respect to the exposure-outcome relationship. RCTs are well-suited to address whether an exposure causes an outcome, because participants in all exposure-level groups are theoretically "exchangeable," that is, the exposure assignment mechanism is known and the characteristics of the participants are the same to the degree that the results would remain if the exposure assignments were switched between groups. The exchangeability allows the difference in the outcomes at each level of exposure to be attributable to the exposure (Hernan 2004). However, RCTs are often unethical or prohibitively expensive and, in the case of exposures such as air pollution, impossible to conduct. Even when RCTs are possible, the results are not always generalizable to a practical target population or a realistic setting in which the particular exposure (treatment) usually is experienced.

Observational studies, in which cohorts of participants with differing exposure histories are followed through time either prospectively or retrospectively often are more practical to conduct. Such observational studies cannot assume "exchangeability," that is, those who are exposed at any level are very likely to differ from those who are exposed at any other level on important factors that confound and/or modify the relationship between the exposure and the outcome (Hernan and Robins 2006). An additional source of lack of exchangeability in such studies is informative censoring over time—*i.e.,* those for whom we have data may differ from those we cannot, or do not, observe in terms of exposure, risk factors and underlying risk of the outcome under study. Due to this lack of exchangeability, the results from observational studies have traditionally been considered to have limited validity for causal inference based on the observed associations between the exposure and outcome. Throughout this discussion, I use the example of a binary exposure; however, though more complicated, these issues can extend to continuous exposures.

4.2.1 Traditional Regression Methods

Traditional regression methods estimate the association between an exposure and the observed outcome. This association is likely to be biased, because the exposure groups usually are not exchangeable. Traditional methods aim to control for this confounding by conditioning on covariates included in the regression model. This can be expressed as

$$E(Y|A,W_k).$$

The association between the exposure *(A)* and outcome *(Y)* are evaluated conditional on fixing the levels of W_k (a vector of confounders and modifiers). If the regression model is properly specified, a condition which is difficult to test and unlikely to hold, then such a model would provide for causal inference within strata of W_k. However, for such an interpretation, important assumptions need to be addressed. First, one assumes there is no unmeasured confounding. This assumption is universal and necessary for any causal inference. Second, one must assume the model is not *estimating* a parameter beyond the scope of the data. This extrapolation of the data is also referred to in the causal inference literature as the experimental treatment assignment (ETA) or positivity assumption. This assumption is generally not addressed in traditional regression. An additional assumption of time-ordering must be made. The covariates and exposure must precede the outcome in order for a relationship to be causal. Finally, the model must be specified correctly. This assumption is

seldom met in traditional regression methods. The model imposes an arbitrary functional form to the data distribution and can cause bias if the model is incorrect.

4.2.2 Counterfactual framework

In epidemiological studies, at any given time, only one exposure and the corresponding outcome are observed for each subject. For example, one subject may be exposed to a high level of air pollution during pregnancy and have a low spirometric parameter. However, one does not know what the spirometric measure would have been if the subject had been exposed to a low level of air pollution. This can be considered a missing data problem because not all possible treatment/outcome combinations are measured in each subject. In contrast, an ideal experiment for the investigation of a causal effect, each subject would receive high *and* low exposure *all else being the same*, and the outcome under each exposure would be observed. This hypothetical set of possible outcomes has been called "counterfactuals" and is referred to the "full" data as opposed to the observed data.

Such an experiment would permit a direct estimation of the causal effect of the treatment of interest (*i.e.*, the difference between the outcomes at each exposure level in each subject under exactly the same conditions), because the only difference between various observations for each person would be the exposure-level. Causal inference methods have been developed to use observational data in this counterfactual framework. These statistical methods allow us to mimic the situation had we been able to observe the outcome in each subject under each possible exposure regimens, all else being held constant in each subject. In this literature, exposures are referred to as "treatments" and I will do the same from here forward. Moreover, I use the word "treatment" to refer to "high exposure" to prenatal air pollution and/or traffic exposure to be defined below.

4.2.3 Marginal Estimates for Causal Inference

The notation used throughout has been adopted in the literature for point-treatment analyses. Extensions to the longitudinal setting can be found elsewhere (Bodnar, Davidian et al. 2004; Brotman, Klebanoff et al. 2008).Throughout the causal inference literature, "**Y**" is used to define the outcome, "**A**" represents the set of all treatments (1=high/0=low exposure in my analyses), and "**W**" is the vector of covariates. For my example, I will use the mean risk difference as the parameter of interest. The following notation defines the counterfactuals of interest: Y_1 is the outcome had all individuals been treated and Y_0 is the outcome had all individuals not been treated. The observed data are only one of the two counterfactual outcomes for each subject because we observe only the outcome having been exposed to *either* high or low exposure, but not both. Causal inference methods use the framework of counterfactuals to approach the estimation as a missing data problem. In contrast to the traditional model above, the parameter of interest for causal inference methods is the counterfactual outcome had everyone received the exposure, $E(Y_a)$ as opposed to everyone who actually received the outcome, $E(Y|A)$. My goal, then, is to model, at the population level, the marginal causal association between the exposure (A) and the outcome (Y) averaged over the covariates (W), which under assumptions discussed below can be written as:

$$E(Y_1)-E(Y_0) = E_w\{E(Y|A=1,W)-E(Y|A=0,W)\}.$$

4.2.4 Assumptions

Causal inference methods rely on five main assumptions: the experimental treatment assignment (ETA) or positivity assumption, no unmeasured confounding, consistency, temporal ordering and correct model specification. These assumptions are not unique to causal inference methods, though are more often made explicit in causal inference literature. The ETA assumption requires that the treatment is not deterministic based on covariates. The probability of treatment for any given treatment mechanism cannot be 0 or 1. In other words, all the treatment levels must be observed for every covariate pattern for the variables included in the final treatment model. As mentioned above, failure to meet this assumption is extrapolation beyond the scope of the observed data and can lead to biased estimation.

The assumption of "no unmeasured confounding" is not testable, unlike the ETA assumption, yet is a critical assumption to all estimations of association. The no unobserved confounding assumption requires that all confounding factors are measured and included in the modeling steps. The consistency assumption requires that the observed outcome is a member of the set of all possible outcomes. The consistency assumption is also not testable, but is a natural consequence of observational data collection, and therefore generally accepted.

Temporal ordering must be maintained; that is, the confounders must precede treatment, which must precede the outcome. The use of directed acyclic graphs is often useful to determine whether this assumption is met. The assumptions for model specification range depending on the estimator of choice and machine learning algorithms can be used to increase the likelihood that this assumption is met. This will be discussed below in more detail.

To obtain an estimate of the marginal causal association, I will discuss three estimators that can be used: inverse probability of treatment weighting and G-computation, which have been building blocks for the final estimator of choice in this dissertation, targeted maximum likelihood estimation.

4.2.5 Inverse probability of treatment weighting

In current epidemiologic literature, the most commonly implemented estimator used for marginal structural models (MSM) is the inverse-probability of treatment weight (IPTW), which reweights the observed population such that the distribution of covariate patterns reflect what would be observed in an RCT. The reweighting has the effect of approximating exchangeability between the exposed and unexposed subjects. It removes confounding through a two-step procedure. The first step requires the creation of weights which are the inverse of the conditional probability of the observed treatment on relevant covariates. This step, referred to as the treatment model step, may be estimated non-parametrically in simple cases with few covariates (by using cell counts) or in the case of a binary treatment with many covariates, may be estimated by a logistic regression. The estimation of weights for continuous treatments is beyond the scope of this work.

Treatment weights defined as the inverse of the conditional probability give the most weight to the observations that are the most unusual (where "usual" is defined by the ideal experiment where A is randomized), hence the term "inverse probability of treatment weights (Mortimer, Neugebauer et al. 2005)." This unstabilized weight can produce large variability in the weights, therefore, it is preferred to use stabilized weights (marginal probability / conditional probability) or $P(A)/P(A \mid W)$, because they provide estimates that can be more

efficient than those obtained by weights that are simply the inverse of the conditional probabilities *(1/P(A | W))*(Robins, Hernan et al. 2000).

For the IPTW estimator, it is essential that the treatment mechanism is correctly specified. The first step is to develop a set of candidate variables that may be associated with the exposure ("treatment") based on previous studies. It is more efficient to include variables associated with the outcome as those are the potential confounders (Lefebvre, Delaney et al. 2008). One can use a data-adaptive algorithm such as the Deletion/Substitution/Addition (D/S/A) algorithm (version 3.1.1) in R (version 2.10.1) to identify the model that best predicted the treatment mechanism(van der Laan). The D/S/A routine is based on cross-validation and the L2 loss function that can be used to predict the conditional expectation of an outcome given a set of explanatory variables (Sinisi and van der Laan 2004). The D/S/A algorithm performs an extensive search of all candidate variables aimed at minimizing the empirical risk function overall index sets of a given size (Sinisi and van der Laan 2004). In a given D/S/A call, the data are split into two groups to model a portion of the data and test it on the remainder. During this process, models that are influenced by outliers will not hold-up in the cross-validation process and will not be chosen as the "best model" (Sinisi, Neugebauer et al. 2006). One can specify restrictions on the size of the model, the order of interactions, the sum of powers, the proportion to split and fold, and the p-value cut-offs. This model selection technique, unlike Akaike's information criterion (AIC), allows for comparison of models with different number of observations, which is particularly useful given the amount of missing data many observational settings.

Once the treatment model is specified, a logistic regression model is fit for the probability of observed high exposure given factors selected from the D/S/A. One can calculate the IPTW with the coefficients from the treatment model. The treatment model is considered a nuisance model because its estimates are only an intermediate step in the full estimation process.

In the second stage, referred to as the structural model step, the outcome is modeled as a function of the treatment, weighted by the IPTW of the first step. The marginal structural model (MSM) coefficients are estimated with an *a priori*-specified model depending on the parameter of interest weighted by the IPTW (*e.g.*, $E(Y_a)=\beta_0+\beta_1 a$).

The standard errors reported from the IPTW weighted regressions are calculated as if the weights are known. If data-adaptive (machine learning) algorithms, such as the D/S/A, are used to obtain this treatment model, the weights are not known. Therefore, it is necessary to recalculate the standard errors to obtain proper confidence intervals for inference on the MSM estimate. Ideally, to estimate the variance of an estimator, one would collect many samples of the same size and implement the IPTW estimator on each sample. The "bootstrap" statistical method is can be used to re-sample observations with replacement from the original sample. The variance of the reiterated estimates is used as the variance of the MSM estimator.

4.2.6 G-Computation

G-computation is a likelihood-based estimator and form of imputation. As mentioned above, as investigators in epidemiology, we observe one outcome under one exposure, but do not know the potential outcome had the exposure been different. Using the observed data, G-computation averages over covariates to create the full data and impute the counterfactuals. The full data can then be used to estimate causal associations.

The first step of G-computation is to fit a regression of the outcome on the exposure and relevant covariates using the observed data. This regression, commonly referred to the "Q model" in the literature, is analogous to a traditional regression of *Y* on *A* and *W*. A machine learning algorithm can be used to fit the Q model, even though it is not appropriate for use with traditional regression (van der Laan, Polley et al. 2007).

The next step is to use the Q model to predict counterfactual outcomes for each observation under each exposure regimen *(Y_a)*. For a binary exposure, as mentioned above, the outcomes are Y_1 for those exposed and Y_0 for those unexposed for each individual. The full data then consists of these estimated counterfactual outcomes. One can then compare the calculated expectations of the outcome in a form of a risk difference or one can use the full data to fit a marginal structural model. G-computation is defined below, where $\hat{E}$ is the average over the empirical distribution of the covariates, *W*.

$$\hat{\theta}^{G-comp} = \hat{E}_w\{\hat{E}[Y \mid A=1,W] - \hat{E}\{Y \mid A=0,W]\} = \hat{E}_w[\hat{Q}^0(W,1) - \hat{Q}^0(W,0)]$$

Although it is simpler to explain these concepts with a binary exposure, there are other options besides resorting to using a continuous exposure, which is more complicated with these methods. See Chapter 6 for an example of quartiles of exposure used to better capture the distribution of the exposure.

4.2.7 Targeted Maximum Likelihood Estimation

The targeted maximum likelihood estimation (TMLE) approach is doubly robust against model misspecification (*i.e.,* the estimator will produce unbiased estimates if either the treatment or the outcome mechanism is modeled correctly). TMLE reduces both the bias and the variance of the targeted parameter. TMLE is maximally efficient and asymptotically unbiased when the nuisance parameters are correctly specified (Gruber 2009).

TMLE requires the modeling of two nuisance parameters before calculating the final parameter of interest. The first is referred to as the Q or outcome model, which is the same model fit for G-computation:

$$E(Y|A,W) \text{ modeled as } Q^0(W, A).$$

The second nuisance parameter, *g,* is analogous to the treatment mechanism of the IPTW estimator and can be defined as:

$$g(A,W) = P(A \mid W).$$

In this example, I employ a binary treatment. Both of these nuisance models can be selected with data-adaptive algorithms such as the Deletion Substitution Addition (D/S/A) algorithm, as described in detail above.

Once $g(A,W)$ is specified, a logistic regression model is fit to obtain the probability of being treated given factors selected. The so-called "clever covariate" is calculated, *h*, with coefficients from this treatment model:

$$h(A,W) = I(A=1)/(g(1 \mid W) - I(A=0)/g(0 \mid W).$$

The clever covariate, *h*, serves as a weight to "target" the parameter of interest in the final step. The clever covariate is added to the Q model to adjust the parameter of interest by incorporating the treatment mechanism and produce an updated Q^1 model and $\hat{\varepsilon}$ is estimated using logistic regression with the logit[Q^0(A,W)] as an offset:

$$\hat{Q}^1(W,A) = \hat{Q}^0 + \hat{\varepsilon} * h(A,W).$$

The final step is to calculate the difference in mean outcomes, given their covariates where

A = 1 and A = 0.

$$\hat{\theta}^{TMLE} = \hat{E}[\hat{Q}^1(W,1) - \hat{Q}^1(W,0)]$$

4.2.8 Population Intervention Models

Population intervention models (PIM) are an extension of the causal inference methods described above. Instead of a counterfactual comparison, a PIM compares the mean outcome in a population under one treatment-specific counterfactual to the mean outcome in the observed population: $E[Y] - E[Y_d] = E(Y) - E_w(Y \mid A = a, W)$ under assumptions. This method accounts for not only the strength of the association of the exposure, but also both the prevalence of the factor in the study population and the non-parametric estimation of the parameter of interest (Hubbard and Laan 2008). Although the PIM will not be the primary parameter of interest in this dissertation, it is mentioned as an option in defining causal parameters.

4.2.9 Influence Curve

As with the IPTW estimator, the standard errors are not correct and can be "bootstrapped" as described above. Statistical inference can also be derived from the influence curve.

$$IC(Y,W \mid \hat{Q},\hat{g},\hat{\theta}) = \frac{I(A=1)}{\hat{g}(1 \mid W)}(Y - \hat{Q}(1,W)) - \frac{I(A=0)}{\hat{g}(0 \mid W)}(Y - \hat{Q}(0,W)) + \hat{Q}(1,W) - \hat{Q}(0,W) - \hat{\theta}$$

The influence curve allows one to calculate how much each observation deviates from the estimate in order to infer the variance of the estimator (Gruber 2009).

4.2.10. Summary

With the use of traditional regression methods, one assumes that the model used is correctly specified. However, if that is not true, the functional form of the chosen model introduces bias. The IPTW estimator requires that the treatment model is correctly specified and the G-computation estimator depends on the outcome model being correct. The TML estimator, on the other hand, will be asymptotically unbiased if either the treatment model or the outcome model is correctly specified, allowing for a more realistic assumption.

The choice of which estimator to use depends on the parameter of interest and the structure of the data. Although not demonstrated in this dissertation, these estimators can be used in conjunction with a parametric model (rather than a risk difference), referred to as Marginal Structural Models (MSMs).

In summary, G-computation allows the estimation of a simple measure of association that assumes no particular model for the regression. In this example, this semi-parametric risk difference is not just a byproduct from an arbitrary regression model. TMLE augments this estimator to be more robust by adding information about the treatment mechanism. PIM is a further extension of these methods, which compares the estimate to the prevalence of the outcome in the study population. These methods allow use of machine learning algorithms to optimize fit and reduce bias. Empirical standard errors are available because the estimate is equivalent to a simple estimating equation approach.

Chapter 5: The Association Between Ambient Air Pollution During Pregnancy and Repeated Measures of Pulmonary Function in Children with Asthma

5.1 Background

Acute and chronic exposure to ambient air pollution has been associated with mortality and respiratory morbidity in many studies (Brunekreef and Holgate 2002; Katsouyanni 2003; Pope and Dockery 2006; Schlesinger, Kunzli et al. 2006). One frequently used assessment of respiratory morbidity is pulmonary function testing (PFT), described in more detail below and in Chapter 2. Some spirometric parameters have been shown to be useful predictors of cardio-respiratory morbidity and mortality (Sin and Man 2005). Associations between acute exposure to ambient air pollutants and pulmonary function in children also have been demonstrated, especially in children with asthma (Mortimer, Tager et al. 2000; Delfino 2002; Mortimer, Neas et al. 2002; Delfino, Quintana et al. 2004).

Fewer studies have examined the effects of ambient air pollution during pregnancy on later respiratory health outcomes (Mortimer, Neugebauer et al. 2008; Turnovska and Marinov 2009). The question of the specific prenatal period that is critical for the long-term effect of air pollution on subsequent levels of lung function in children is not known. Ambient air pollution exposures in pregnant women have been shown to lead to adverse early-life events, most commonly low birth weight (Ritz and Yu 1999; Bobak 2000; Ritz, Yu et al. 2000; Maisonet, Bush et al. 2001; Maroziene and Grazuleviciene 2002; Ritz, Yu et al. 2002; Lee, Ha et al. 2003; Woodruff, Parker et al. 2003; Ritz, Wilhelm et al. 2007; Ritz and Wilhelm 2008; Wang, Ding et al. 1997), preterm birth (Liu, Krewski et al. 2003; Wilhelm and Ritz 2005; Dugandzic, Dodds et al. 2006; Huynh, Woodruff et al. 2006; Leem, Kaplan et al. 2006; Jalaludin, Mannes et al. 2007), small for gestational age (Parker, Woodruff et al. 2005) and respiratory morbidity and mortality in the first year of life (Lacasana, Esplugues et al. 2005; Woodruff, Parker et al. 2006). These studies suggest that these exposures influence lung function in a manner similar to that which has observed for maternal smoking during pregnancy (Cunningham, Dockery et al. 1994; Gilliland, Berhane et al. 2000; Li, Gilliland et al. 2000). Pregnancy is a critical time where ambient air pollution may influence cell proliferation and organ development (Sorlie, Kannel et al. 1989; Barker 2004). As mentioned above (Chapter 2.3), the lungs begin to develop at 6 weeks gestation and airways are in place by 26 weeks gestation and this period may be a critical period for exposure.

Recent evidence from the Fresno Asthmatic Children's Environment Study – Lifetime Exposures (FACES-LITE) study suggests that higher levels of exposure to NO_2 during the second trimester, CO during the entire prenatal period, and PM_{10} during the first trimester, may be associated with decrements in baseline pulmonary function among children with asthma at the ages of 6-11 years old (Mortimer, Neugebauer et al. 2008). I hypothesized that the effect may continue throughout childhood. I tested this hypothesis through examination of the effect of the prenatal exposures to ambient air pollutants on repeated measures of pulmonary function observed over a four-year follow-up of the FACES-LITE study population.

The aim of this analysis is to estimate the marginal (population-level) causal association for each pollutant during each time period of pregnancy across a range of spirometric parameters that reflect different aspects of lung function. In contrast to previous studies that

have been limited to analyses of conditional associations, I will employ recently developed causal inference estimators that allow analysis of observational data in a framework analogous to that of randomized controlled trials. In particular, I will implement targeted maximum likelihood estimation (TMLE) to control for confounding yet allow a marginal interpretation. As discussed further in Chapter 4, the TMLE approach is doubly robust against model misspecification, and reduces bias and variance of the targeted parameter. TMLE is maximally efficient and asymptotically unbiased when the nuisance parameters are correctly specified.

I will implement TMLE to estimate the causal association between prenatal exposure and ambient air pollutants during each pregnancy period (each trimester and the entire pregnancy) and six pulmonary function tests, measured every 6 months for 4 years, with children aged 6-11 at baseline. As mentioned above (in Chapter 4), in the causal inference literature, exposures are referred to as "treatments." The treatments in these analyses are a series of metrics that characterize prenatal exposure to ambient air pollutants including carbon monoxide (CO), nitrogen dioxide (NO_2), ozone (O_3), and particulate matter with an aerodynamic diameter smaller than 10μm (PM_{10}). For each analysis, a single pollutant is summarized both over a single trimester and for the entire pregnancy.

5.2 Methods

5.2.1 Study Population

Data for this study were collected within the context of the Fresno Asthmatic Children's Environment Study (FACES), a longitudinal epidemiologic study of 315 asthmatic children in Fresno, California. FACES-LITE, a separately funded sub-study, includes 232 members of the cohort for whom prenatal or early life exposure could be obtained. I restricted the sample of observations to those who were 13 years old or younger since few participants were followed past the age of 13 (n=8).

Children and their families were recruited through community-based advertisements, school nurses and local physicians. Eligibility criteria included physician-diagnosed asthma, age 6-11 years at the time of the first interview, and having a primary residence within 20 kilometers of the California Air Resources Board's Fresno First Street air monitoring site. Only children with asthma who had been symptomatic or used medication or had a physician visit for asthma in the 12 months previous to enrollment were eligible. The protocol consisted of an in-person office visit (baseline), followed by a series of home surveys, telephone calls and office visits. An in-person, baseline questionnaire was administered in English or Spanish. During the baseline interview, an extensive array of participant characteristics was ascertained (Table 1).

5.2.2 Outcome Ascertainment

Spirometric parameters for these analyses were obtained from force expiratory maneuvers performed with a dry, rolling-seal spirometer (Spiroflow; P.K. Morgan Instruments, Andover, Massachusetts.) During the in-person field office visits, participants were in the seated position with nose clips. Up to eight attempts were allowed to obtain up to three acceptable tracings as described previously (Mortimer, Fallot et al. 2003). Although pre-bronchodilator and post-bronchodilator maneuvers were obtained, this analysis is restricted to pre-bronchodilator measurements since pre-bronchodilator function is more reflective of the usual state of the subject's lung function. Time-volume and flow-volume curves were reviewed

on screen by two investigators to judge whether the tracings were acceptable for analysis. Tracings were acceptable if they lasted more than 2 seconds and the flow/volume curve reached a plateau. The mean of the first acceptable three (but not less than two) acceptable tracings was calculated (Mortimer, Fallot et al. 2003).

The spirometric parameters reflect several aspects of pulmonary function. Volume measures including forced vital capacity (FVC) and forced expiratory volume in one second (FEV_1) were used because they are commonly used spirometric parameters. Flow measures including forced expiratory flow between 25 and 75% of vital capacity (FEF_{25-75}), and forced expiratory flow at 75% of vital capacity (FEF_{75}) were examined because they reflect the physiology of asthma. Also considered were the ratio of FEV_1 to FVC (FEV_1/FVC), and the ratio of FEF_{25-75} to FVC (FEF_{25-75}/FVC). The ratio of FEF_{25-75}/FVC is among the most interpretable physiologically.

Many studies have demonstrated a reduction in forced expiratory *flows* in infants exposed to parental smoking (Hanrahan, Tager et al. 1992; Stocks and Dezateux 2003). In the FACES-LITE baseline study, spirometry measures of both *flow* (in the case of CO and PM_{10}) and *volume* (for NO_2) were affected by the pollutants therefore, I chose to look at both the flow, volume and ratio measures to evaluate the effects of prenatal exposures on the repeated measurements of pulmonary function. Ambient air pollutants penetrate the small airways, the sites of pathology in asthma. The following spirometric parameters are measures of the small airways: FEF_{25-75}, FEF_{75} and FEF_{25-75}/FVC.

5.2.3 Exposure Assessment

At the 6-month office visit, parents were asked to report the street address, city and state of all residences at which the mother lived during pregnancy as well as homes at which the child lived prior to enrollment in the study. Only addresses at which they lived for at least three months were recorded. Each address was geo-coded with “EZ Locate Client" v.1.61, by Tele Atlas.

Pollutant concentrations were obtained from the Aerometric Information Retrieval System database supported by the U.S. Environmental Protection Agency. Centrally located monitors sampled several pollutants including NO_2, CO, PM_{10} andO_3. Pollutant metrics were determined by time-averaging the raw database concentrations. The estimates were mapped to the residences or ZIP code centroids based on inverse distance weighting of the monthly average concentrations from the air monitoring stations (up to three) located closest to the residence location.

Residences were assigned a quality code, based on the distance to the nearest monitor(s).

Quality Code	Distance between residence and nearest monitor
1	< = 5 km
2	5 - 50 km for NO_2, O_3, and PM_{10} 5 - 25 km for CO
3*	50 - 100 km for NO_2, O_3, and PM_{10} 25 - 50 km for CO

*Code 3 addresses were not included in the calculation of the exposure metrics.

Pollutant levels for the years for which members of this cohort were *in utero* (1989-1997) and throughout enrollment in FACES were obtained. Twenty-four hour averages were calculated for NO_2 and PM_{10} and 8-hour maxima were calculated for CO and O_3. These metrics were chosen because they correspond to time intervals used for regulatory standards. For the prenatal analyses described in this report, these monthly values were averaged once across the entire pregnancy and separately for each trimester.

I hypothesized that the magnitude of the effect of prenatal exposures may differ as the child grows. Therefore, I tested whether the effect, if it exists, differs across age groups. I coded ages into three categories (6-8, 9-10, 11-13 years) based on the relative homogeneity of pulmonary function values within groups and to balance the sample sizes in each group. I then performed extensive analyses on the age-stratified data, as outlined below. The study protocol was approved by the Committee for the Protection of Human Subjects of the University of California, at Berkeley. Written informed consent was obtained from parents/legal guardians prior to enrollment.

5.2.4 Statistical Methods

In a point treatment study, where there typically is a measure of exposure and a measure of an outcome, a marginal estimate of the effect of an exposure on a population can be obtained. In a longitudinal study with repeated measures of exposure and outcomes, marginal structural models (MSMs) and targeted maximum likelihood estimation (TMLE) can account for time-dependent confounding and factors on the causal pathway, as well as provide a marginal interpretation of the causal effect (Ko, Hogan et al. 2003; Bryan, Yu et al. 2004; Brotman, Klebanoff et al. 2008).

The analysis described below implements a less common study design. Separately over a series of pulmonary function outcomes, I looked at the effect of a single treatment, exposure of the mother to air pollution, on pulmonary function at a series of ages (range: 6-13 years). Essentially, this is a point treatment exposure with repeated measures of the outcome.

As mentioned in Chapter 4, TMLE is based on counterfactual theory. Briefly, counterfactuals are the set of possible outcomes that would be observed under each possible treatment, if, contrary to fact, each person could be observed after exposure to each level of the treatment. The goal of this analysis is to estimate, at the population level, the average causal association between prenatal air pollutants and lung function parameters, which can be written as

$$E(Y_1)\text{-}E(Y_0) = E_w\{E(Y|A=1,W)\text{-}E(Y|A=0,W)\}.$$

under assumptions, where $E(Y_1)$ and $E(Y_0)$ are the population mean lung function outcomes when all individuals in the population have high and low exposure, respectively.

As noted in Chapter 4, TMLE requires the modeling of two nuisance parameters before calculating the final parameter of interest. The first is referred to as the Q or outcome model and is identical to G-computation:

$E(Y|A,W)$ modeled as $Q^0(W, A)$.

The second nuisance parameter, *g*, is analogous to the treatment mechanism of the inverse probability of treatment weighting (IPTW) estimator and in the case where A is binary, can be defined as:

$$g(A,W) = P(A \mid W).$$

The most straightforward way to employ a TML estimator is with a binary treatment. Therefore, I dichotomized the exposure to each pollutant metric into high and low with a cut-point at the median. Both of these nuisance models were selected by the Deletion Substitution Addition (D/S/A) algorithm.

Once *g(A,W)* was specified, I fit a logistic regression model for the probability of high prenatal pollutant exposure given variables selected from the D/S/A. I calculated the clever covariate, *h*, with coefficients from this treatment model:

$$h(A,W) = I(A=1)/(g(1 \mid W) - I(A=0)/g(0 \mid W).$$

The clever covariate, *h*, serves as a weight to "target" the parameter of interest in the final model. The clever covariate is added to the Q^0 model to adjust the parameter of interest by incorporating the treatment mechanism and produce and updated Q^1 model and $\hat{\varepsilon}$ is estimated using logistic regression with the logit[Q^0(A,W)] as an offset:

$$\hat{Q}^1(W,A) = \hat{Q}^0 + \hat{\varepsilon} * h(A,W).$$

Finally, I calculated the difference in mean pulmonary function test, given their covariates where A = 1 and A = 0.

$$\hat{\theta}^{TMLE} = \hat{E}_w[\hat{Q}^1(W,1) - \hat{Q}^1(W,0)]$$

I specified that the models should be restricted to a size of 10, second order interactions, and a maximum sum of powers of 3—*i.e.*, up to 10 variables in the model, two-way interaction terms between any of the variables, all the variables up to second order polynomials, and p-value cut-offs of 0.05, 0.2, 0.6 and 1.0 (Mortimer, Neugebauer et al. 2008). When I specified a series of p-value cut-offs the resulting set of iterations contained only variables that were significantly associated with the treatment, at the prescribed p-value (<0.05, <0.2, <0.06) and one set of iterations that included all the variables (*i.e.*, the default of $p<1.0$). This allowed for varying stringency in which variables I considered for the model. I specified the identifying unit for each individual in the data for both the *g* and *Q* models, so that the repeated measures of each individual would be linked and treated accordingly. This model selection technique, unlike Akaike's Information Criterion, allows for comparison of models with different number of observations, which is particularly useful given the amount of missing data in this study.

Data were collected on numerous demographic and environmental characteristics for all the children and parents in the FACES study. However, the candidate variables for each of the nuisance models need to respect the time ordering, that is, they must precede the treatment (prenatal exposure to pollutants.) These characteristics are listed in Table 5.5. There are two variables that were included as surrogates for prenatal socioeconomic status even though they violate time ordering: low family income (greater than $30,000/year at baseline) and home ownership at baseline. The assumption was made that these values are relatively stable over time, and; therefore, measures collected at ages 6-11 years reflect relative ranking of income and home ownership at the time of birth.

I calculated the standard errors to obtain proper confidence intervals for inference on the TMLE estimate with an influence curve:

$$IC(Y,W \mid \hat{Q},\hat{g},\hat{\theta}) = \frac{I(A=1)}{\hat{g}(1 \mid W)}(Y - \hat{Q}(1,W)) - \frac{I(A=0)}{\hat{g}(0 \mid W)}(Y - \hat{Q}(0,W)) + \hat{Q}(1,W) - \hat{Q}(0,W) - \hat{\theta}$$

The influence curve allows one to calculate how much each observation deviates from the estimate in order to infer the variance of that estimator. I calculated confidence intervals at the 95%, 90% and 80% level. As with the D/S/A, I specified the identifying unit for each individual, so that the observations would be interpreted as repeated measures of each individual rather than independent observations. Analyses were performed using R software (R Foundation for Statistical Computing, Vienna Austria, version 2.10.1, package: DSA, version 3.1.1).

5.3 Results

Of the 315 members of the FACES cohort, 232 subjects were in the FACES-LITE cohort. Those for whom prenatal or lifetime exposure was not available were excluded (N=909). For this analysis, there were 162 members of the cohort for whom prenatal exposure to CO, NO_2, O_3 and PM_{10} and spirometric parameters could be obtained. The characteristics of this subset of the population are described in Table 1. As in the Fresno area, the majority of this study population is either white or Hispanic. There are slightly more males (58%) than females. At baseline, approximately one fifth of the study population had severe asthma, about half had moderate asthma and one third had mild asthma. Residential history was provided for an average 87% of the pregnancy. The study population did not differ between those used in this analysis and those in the original FACES-LITE cohort.

This study population consisted of 162 children contributing 947 observations of pulmonary function. An additional 909 observations were not included because there were missing data on the spirometric parameters. The table below lists the number and reason for missing the spirometric parameters.

Number of Observations Missing (n=909)	Reason for Missing Spirometric Parameters
560	particular visit not due by the time data collection ended
158	missed due to drop out
100	missed visits
58	spirometric errors
33	unacceptable tracings

The distribution of each of the pollutants for the entire pregnancy and each trimester are plotted in Figures 1-4 for each of the pollutants. The entire pregnancy (labeled as "4") has the smallest variance due to spanning over multiple seasons compared with the individual trimesters.

Figures 5.8 (show the pollutant averages by year of birth. Ambient pollutant levels have been decreasing throughout the 1990's nationwide, and it is apparent in these plots that the same is true for the Fresno area, particularly for PM_{10} (Figure 7). In Figures 5.9-5.12, the plots show the distribution of each pollutant for the entire year compared across each season (Winter: October-January, Spring: February-May, Summer: June-September). The seasons in the Fresno area correspond to time periods characterized by high particle exposure (winter), high oxidant exposure (summer), and a transition period (spring).

The correlations between the pollutant metrics are shown in Table 3. For the entire prenatal period, NO_2 and CO are moderately correlated ($\rho = 0.78$). The remaining the pollutants are not highly correlated. O_3 is negatively correlated with NO_2 and CO ($\rho = 0.40$ and $\rho = -0.43$, respectively). The pollutant concentrations in the second trimester showed the highest correlation with concentrations during the entire pregnancy for all pollutants. Table 4 shows the median for each of the pollutants and the number of observations above and below the median in each age group. Although the majority of the children were born in CA (93%), there was a substantial range of pollutant exposures over the entire pregnancy (NO_2: 9.0-48.6 $\mu g/m^3$, CO: 0.4-3.6 $\mu g/m^3$, PM_{10}: 17.9-94.3 $\mu g/m^3$, O_3: 13.9-77.4 $\mu g/m^3$).

5.3.1 Targeted Maximum Likelihood Estimation

Tables 5.5a-d show which of the covariates were selected for the outcome (Q) and treatment (g) models. The full results of the TMLE analysis are presented in Tables 5.6-5.9 for NO_2, CO, PM_{10} and O_3, respectively. The results are stratified by age group (6-8, 9-10, 11-13) and exposure period (entire pregnancy and each trimester).

In Table 5.6a, the most striking results are during the first and second trimesters where nearly all the PFTs across all age groups show a deficit if, contrary to fact, everyone had been exposed to above median concentrations of NO_2 compared to below median. Statistically significant results were concentrated in the ratio measures. For example, among the 6-8 year olds, higher exposure to NO_2 during the second trimester resulted in an age group-level decrease in FEV_1/FVC of 0.07, FEF_{25-75}/FVC of 0.31 and FEF_{75} of 0.35 (44% decrease). All three of these parameters reflect the decrease of airway obstruction, with FEF_{25-75}/FVC being the most interpretable physiologically. Among the older age group (ages 11-13 years) higher exposure to NO_2 during the first trimester resulted in a decrease of 0.06 for FEV_1/FVC and 0.23 for FEF_{25-75}/FVC. In Table 5.5b, all of the spirometry measures, with the exception of FVC, show the percent change increasing as the children age. There are also two results that are unrealistically large among the 6-8 year old group during the second trimester(FEF_{25-75}/FVC and FEF_{75}) with 34% and 44% deficits had everyone been exposed to above versus below median levels of NO_2.

Exposure to carbon monoxide during the second trimester shows deficits in PFTs in all age groups (Table 5.7b). However, FEV_1/FVC was the only spirometric measure that had statistically significant results during the first and second trimester as well as the entire pregnancy. The second trimester was the only time period that showed a percent change increase with age. Among 9-10 year olds, FEV_1/FVC was 0.03 lower had everyone been exposed to above median levels of CO during their second trimester compared to had everyone been exposed to below median levels. Additionally, high CO exposure during the entire pregnancy resulted in lower pulmonary function tests among the youngest (6-8 years) and oldest age group (11-12 years). Compared to NO_2 and CO, there was less consistency in the results for PM_{10} and O_3 (Table 5.8 and 5.9).

As mentioned earlier, I only included covariates which respected the time ordering. The temporal ordering assumption was met because the covariates in the treatment model preceded the prenatal exposure, which in turn, preceded the pulmonary function measurements in childhood. The "no unobserved confounding" assumption is not one that can be tested empirically. Since time activity data were not collected, it is unknown how much time the mother spent in or near the home or to the extent to which she spent time in

locations with ambient air pollution conditions that were substantially different from those estimated based on residence. Time activity data were not available to address this issue. Another important potential confounder or effect modifier is smoking in the home at the time of pregnancy. Although this could be defined in the main FACES study, smoking in the home was not obtained for the prenatal period, other than maternal smoking which was rare. It is possible that the socioeconomic status during pregnancy may have been different from the socioeconomic status at the baseline of the study when the child was between 6 and 11 years of age. Although this study has extensive covariate data compared to some other studies, many of the covariates included were proxies for potential confounders and there may be misclassification of those factors.

The consistency assumption is made, implying that the observed outcomes are a subset of all of the possible outcomes. The models, chosen by the D/S/A algorithm are assumed to be correct. However, it is only required that one of the two nuisance models is correct (*i.e.,* it allows for error in the model specification for either the *g* or *Q* model.) More details about the assumptions can be found in Chapter 4.

5.3.2. Traditional Regression Estimation

Results from the traditional analyses are in Tables 5.6c-5.9c. In general, the results tend to be of greater magnitude and demonstrate a larger number of "statistically significant" results, however, they cannot be directly compared for two reasons. First, they are conditional associations (stratified by all potential confounders in Table 5.5a) and do not have a marginal (population level) interpretation. Additionally, the method of calculating the statistical inference is not the same as that of the causal inference methods described above.

5.4 Discussion

As mentioned above, Mortimer *et al.* found, in the same cohort children, that increased 24-hour average NO_2 during the second trimester was associated with lower FVC and FEV_1 at baseline examination (mean age = 8.56, SD = 1.7, range 6-11) (Mortimer, Neugebauer et al. 2008). In addition, higher 8-hour maxima of CO during the entire pregnancy were associated with deficits in FEF_{25-75}. The best predictors of the three remaining spirometric parameters (FEV_1/FVC, FEF_{25-75}/FVC and FEF_{75}) were early life exposure to air pollution in the years of childhood prior to enrollment in the FACES study. These baseline results gave a starting point to examine the follow-up period of this cohort, which provided us with repeated measurements made on the same cohort of children as they aged.

The results from the FACES-LITE baseline paper (Mortimer, Neugebauer et al. 2008) cannot be directly compared to those of the current analysis for several reasons. First, the baseline analysis estimated a conditional association, and the current analysis provides a marginal association. Second, the metric was treated as a continuous variable and contrasted by an inter-quartile range increase, and the current analysis used a dichotomized exposure. Finally, and most importantly, the baseline analysis described what factors best predict each pulmonary function measure and factors included additional metrics such as lifetime exposure during the first 3 and first 6 years of life, which are not used in the current analysis. Nevertheless, these independent analyses both support the hypothesis that prenatal exposures to ambient air pollutants during the first and second trimesters can have detrimental effects on pulmonary function of children with asthma.

The causal associations of the current analysis differ by age group. The percent change in spirometric measures increased as the children aged had everyone been exposure to high NO_2 in the first trimester or high CO in the second trimester. In some cases, a greater and more consistent effect is seen in the older children. It is possible that as children grow, the effects of prenatal ambient air pollution become more apparent during times of rapid lung growth and development, or may not be seen until after puberty, yet still affecting pulmonary function in the long term. Although this analysis does not capture the tracking of lung function, it is well established that children have a strong tendency to maintain their relative levels of spirometric performance through childhood into adulthood (Rasmussen, Taylor et al. 2002). Studies have found evidence that deficits in lung function persist and track over time (Dockery, Ware et al. 1985; Wang, Dockery et al. 1993). It has been hypothesized that airway remodeling in childhood can lead to respiratory outcomes in adulthood (Rasmussen, Taylor et al. 2002).

There are other potential explanations as well. Older children may have higher quality spirometry and may provide a more precise estimate due to smaller measurement error. The younger children did take more medications, both for prevention and rescue of asthma. This cohort effect coupled with younger children being less exposed could influence the result. Due to time-ordering restrictions, medication use was not a candidate for selection by the D/S/A into the nuisance models.

It would be interesting to examine the effect of prenatal exposure on *growth* of pulmonary function over time, rather than the level of pulmonary function at a given time. Although I considered using the change in spirometric measures between visits to examine growth, I decided against it for several reasons. First of all, because the data are fraught with missing values, using a change score would censor many subjects. Second, although pulmonary function growth curves can be very informative of the overall pattern of growth, in this age range, the considerable changes between visits would require too much interpolation since the data, at best, is collected every six months. Lastly, a change score would inherently adjust for the previous spirometric measure. However, previous pulmonary function is not a confounder because it is on the causal pathway from prenatal air pollution exposure to the indexed spirometric measure and adjusting for it could bias the effect estimates (Hernan, Hernandez-Diaz et al. 2004).

Many of the significant results across all pollutants and time periods were for the pulmonary function test, FEV_1/FVC. This ratio is the percentage of vital capacity which is expelled in the first second of maximal expiration. Decreases in this ratio are a sign of obstructive airway disease. As Ramsey found in a study of asthmatic children in Connecticut, FEV_1/FVC is useful for distinguishing asthma severity (Ramsey, Celedon et al. 2005).

The most notable results were found for NO_2 where the effects were most consistent for exposure during the first and second trimester. An important consideration in the determination of the effects of various environmental exposures on respiratory health in children is the state of development of the lungs and the immune system at the time of exposure. This most critical period with respect to spirometric parameters in this study is during these first two trimesters when the airway is developing. The lungs begin to develop at 6 weeks of gestation and continue through distinct phases of progression. The airways and blood vessels are in place by 26 weeks' gestation.

The results from the third trimester were quite different from the other trimesters in the NO_2 results. The third trimester is a problematic time in epidemiological studies due to the various lengths of pregnancies and the way in which trimester dates are calculated. The length of third trimesters varies tremendously due to premature births, which makes it difficult to assign exposures and compare across groups with different gestational ages. On the other hand, in this cohort, there are few preterm births (n=16, 10%) thus it should not have influenced the results considerably. It is possible there is no association, as the results suggest, between NO_2 exposure during the thirds trimester and lung function in childhood.

The impact of prenatal exposure to CO on pulmonary function was concentrated in the second trimester and effects were significant only among the two younger age groups for one pulmonary function test, FEV_1/FVC. Due to the change in pollution profile over time, the distribution of above and below the median was not consistent across age groups (see Table 5.4). The oldest age group had higher exposure and the younger were less exposed. In addition, there were fewer children in this age group, so the lack of balance may have had a bigger impact on this age group which lead to inconsistency in the results for the children aged 11-13.

The inconsistent results for prenatal PM_{10} exposure on pulmonary function may be explained by the large decrease in ambient PM_{10} exposure seen through the 1990s. Other studies have often found effects of particulate matter; however, many of them measured $PM_{2.5}$. Although that measurement was not available, future studies should look at finer particulate matter. Additionally, the PM_{10} measurements were taken every 6 days rather than hourly for other ambient air pollutants. The values assigned to the pregnancy and first trimester for PM_{10} were based on fewer exposure measurements of daily exposure, therefore the trimester and prenatal estimates of exposure were more variable and potentially less accurate than those of NO_2, CO or O_3 which were measured daily.

The inconsistency of the results of O_3 during pregnancy was not surprising considering O_3 was never chosen as a predictor of deficits in pulmonary function in the baseline models. As shown in Table 5.3, ozone is inversely correlated with the NO_2. If there is an effect of NO_2, one would not expect to find association between O_3 and lung function in this study.

In this study, maternal smoking during pregnancy was rare and second hand smoke exposure during pregnancy was not measured. Smoking in the home at the baseline of the study was measured, but it is not known how representative this variable would be in assessing risk of prenatal exposure to tobacco products. In other studies, *in utero* exposure to tobacco smoke is known to be associated with lower lung function later in childhood (Gilliland, Berhane et al. 2000) and the occurrence of asthma (London, James Gauderman et al. 2001). *In utero* exposures to maternal smoking have been found to be associated with decreased pulmonary function in infants (Tager, Weiss et al. 1983; Moshammer, Hoek et al. 2006) and children of school age, especially for small airway flows (Gilliland, Berhane et al. 2000). Studies of maternal smoking during pregnancy and pulmonary function in children have found associations with deficits of 2-8% in the following spirometric measurements: FEF_{25-75} (Cunningham, Dockery et al. 1994; Cunningham, Dockery et al. 1995; Gilliland, Berhane et al. 2000; Li, Gilliland et al. 2000), FEF_{75} (Cunningham, Dockery et al. 1994; Gilliland, Berhane et al. 2000) and FEV_1/FVC (Cunningham, Dockery et al. 1995; Li, Gilliland et al. 2000).

Previous research on exposure to tobacco smoke provides insight into possible mechanisms of action. *In utero* exposure to tobacco products has been shown to have effects on measures of airway function and not on lung volumes, which suggests that there is a vulnerable period with respect to respiratory tract development in the first and early second trimesters (Hanrahan, Tager et al. 1992). Since tobacco smoke contains many compounds found in ambient air (NO_2, CO, PM_{10}, polycyclic aromatic hydrocarbon), it is reasonable to hypothesize that ambient pollutant exposures also alter the development of the fetal respiratory tract through similar mechanisms. It has been demonstrated that long-term exposures to particulate matter from second hand smoke and ambient air pollution adversely affect biological responses such as pulmonary and systemic oxidative stress (Pope and Dockery 2006). Additionally, long-term exposures to particulate matter have been associated with small airway remodeling (Brauer, Avila-Casado et al. 2001; Churg, Brauer et al. 2003), a sign of obstructive airway disease including asthma.

The FACES study population does present some strengths and limitations. The study was geographically homogeneous, with a majority of births in or near Fresno, which reduces confounding by restriction. In addition, this area is a highly exposed part of the country in terms of air pollution and provides an opportunity to examine effects in an area of non-attainment for multiple pollutants (PM_{10}, O_3). On the other hand, this homogeneity, as mentioned above, may not provide a large enough contrast between high and low exposures to accurately assess the effects of the pollutants. Furthermore, the FACES study was conducted on a highly vulnerable population – children with asthma. This is an important subgroup and is expected to be among those most affected by ambient air pollution. However, the results may not be generalizable to children without asthma. Additionally, it is not known how these associations will persist through puberty and into adulthood.

Though this study had a small sample size (n=162 children), the repeated measures of pulmonary function created more opportunity for increased observations (n=947 spirometry parameters). These findings do bring to attention the need to continue to follow cohorts as they age to measure the effects throughout and beyond puberty where growth accelerates. Larger samples would also allow for identification of additional subgroup differences in populations. I was only able to look at age group differences with the sample size at hand, but it is possible the effects of prenatal ambient air pollution on pulmonary function interact with race, socioeconomic status or other subgroup differences.

Because a majority of the study population was within a relatively close area, in or near Fresno, it is possible the range and level of the pollutants lowered the power to detect an effect of high versus low air pollution. If the difference between high and low was not large enough and there was an effect at a higher contrast of exposure, the high-low difference might not have been detectable. Future studies with a broader geographic range would help determine if this happened, however, the trade-off is the potential increase in confounding by other characteristics. Although the ranges are 5-fold in most cases, the inter-quartile range is quite narrow, and provided a small contrast between above and below median populations (See Table 5.4).

The exposure assessment of air pollution measurements were obtained from the central monitor site in Fresno. This exposure assessment is dependent on the estimation of the central monitor and the distance from the monitor. A limitation of pollutants measured at

central monitors is the lack of spatial resolution necessary to capture both the temporal and spatial variability of pollutants from local-scale traffic. Concentrations may be lower in areas more distant from the monitors, which could cause bias in a study examining an association with a health effect (HEI 2010). Exposure misclassification varies by averaging period of the pollutant and the length of exposure period. For example, in a study of prenatal exposure to the criteria pollutants, PM may be measured every 6 days and CO is measured every hour therefore the precision of the CO measurements will be greater. Additionally, an exposure period of a trimester versus the entire pregnancy will affect the variability of the measurement.

This study did not consider mixtures of pollutants. Each of the pollutants was analyzed in isolation. It is likely that a combination of pollutants rather than one single pollutant is responsible for decrements in pulmonary function and that they act in concert to affect the outcome. However, the relatively small sample size of this study complicated any attempt to disentangle the effects of highly correlated pollutants. Future studies should explore additional methods to analyze multiple pollutants or consider measurement of exposure to sources such as proximity to high traffic areas.

Traditionally, analyses of pulmonary function have included covariates such as: sex, height and weight at time of spirometry, low birth weight or prematurity, current smoker in home in childhood, asthma severity, medication use at time of spirometry, history of atopy and breastfeeding, the pollutant level at interval before pulmonary function test. I did not consider those as confounders for my analysis, however, because their inclusion would violate proper time ordering. Many of these variables, such as low birth weight, could be on the causal pathway; therefore, they could not be confounders. I was interested in the direct and indirect effects of prenatal exposure to air pollution on pulmonary function in childhood. If I had controlled for them I would have only considered the direct effects, not those potentially mediated by other factors. A few of the variables warrant further discussion as noted below.

Height is an important predictor of pulmonary function and is included in most analyses of pulmonary function (1995). However, as noted above, it is possible that height is on the causal pathway if it is influenced by prenatal exposure to air pollution. It could be argued that even if there is no biological influence on height on exposure, height could be an empirical confounder and should be considered. Though, because this would violate time-ordering, however, I did not include it as a candidate. A similar argument can be made for weight. Furthermore, it is unlikely that these characteristics are associated with air pollution. If they were due to some developmental effect, it would be on the causal pathway.

Over the course of the study, the participants took a variety of medications for asthma treatment. Each medication was classified as a rescue or controller medication. Medication may be on the causal pathway between prenatal exposure and pulmonary function. For example, a participant with higher prenatal exposure may take more or different kinds of medication that improve his/her pulmonary function, which makes prenatal exposure appears to be 'protective' and improving pulmonary function. Adjustment for medication use (which occurred after prenatal exposure) would bias the effect estimate.

In much larger datasets, prenatal exposures have been shown to be associated with small but statistically significant decreases in birth weight and gestational age, both of which have been shown to influence pulmonary function later in life. In this case, these birth

outcomes would be on the causal pathway between prenatal exposures and pulmonary function growth and should not be considered as confounders in this analysis.

A potential limitation is the effect of air pollution immediately (days) prior to the pulmonary function test. However, it is unlikely that it would be correlated with prenatal exposure. Another concern is that prenatal exposure may be correlated with childhood exposure (*i.e.*, childhood exposure may be on the causal pathway or an independent risk factor.) As has been shown in the literature of passive smoke exposure, it is difficult to disentangle the effect of prenatal and postnatal exposures due to strong correlation. Despite the fact that I did not account for this potential mis-assignment of the critical period, if it turns out that early life rather than the prenatal period is more critical, the public health recommendations to avoid air pollution for pregnant women and small children and lower air pollution standards on the regulatory level would be similar.

Despite these potential limitations, the findings of this study are important for several reasons. There is an expanding literature on the harmful effects of prenatal exposures on children's health. These data provide support for the hypothesis for the existence of causal associations between prenatal exposures to air pollution and spirometric measures among school-aged children. If these associations remain in adulthood, children with prenatal exposure may be at greater risk for serious and debilitating respiratory disease.

This is the first application of TMLE for research of prenatal air pollution exposure on health outcomes. This novel statistical method requires fewer parametric assumptions and gives a marginal estimate in the presence of confounding with less bias and more efficiency than traditional regression methods. The original estimate, Q, is biased due to the variance-bias trade-off since it estimates the entire density. This model is analogous to the traditional model. The unique property of TMLE is that it "targets" the parameter of interest by adding a clever covariate. This updating step in the model controls for confounding by weighting the exposure to resemble a randomized controlled trial. In addition to reducing bias and using an optimally efficient estimator, this analysis also provides a marginal estimate of the effect of prenatal exposure to air pollution on pulmonary function in childhood. This marginal effect is more appropriate than a conditional estimate for useful inference at the population level of interest.

5.5 List of Tables and Figures

Tables

below median levels of 24-hour averages during pregnancy intervals, based on the targeted maximum likelihood estimation.

5.8b. Same results as above (PM_{10}) presented as percent change in pulmonary function by age group.

5.8c. Traditional method: Conditional association between PM_{10} exposure during pregnancy and mean pulmonary function by age group (conditional on all variables in Table 5.5c).

5.9. Results for O_3: causal inference methods, presented in absolute difference and percent difference, and traditional methods. Deficits in pulmonary function are in **bold**.

5-9a. Causal inference methods: Marginal association between O_3 exposure during pregnancy and mean pulmonary function by age group. Results compare above versus below median levels of 24-hour averages during pregnancy intervals, based on the targeted maximum likelihood estimation.

5.9b. Same results as above (O_3) presented as percent change in pulmonary function by age group.

5.9c. Traditional method: Conditional association between O_3 exposure during pregnancy and mean pulmonary function by age group (conditional on all variables in Table 5.5d).

Figures

5.1. Distribution of NO_2 during each trimester (1, 2, 3) and the entire pregnancy (4).
5.2. Distribution of CO during each trimester (1, 2, 3) and the entire pregnancy (4).
5.3. Distribution of PM_{10} during each trimester (1, 2, 3) and the entire pregnancy (4).
5.4: Distribution of O_3 during each trimester (1, 2, 3) and the entire pregnancy (4).
5.5. Distribution of NO_2 during the entire pregnancy by year of birth.
5.6. Distribution of CO during the entire pregnancy by year of birth.
5.7. Distribution of PM_{10} during the entire pregnancy by year of birth.
5.8. Distribution of O_3 during the entire pregnancy by year of birth.
5.9. Distribution of NO_2 for the whole year and by season of birth.
5.10. Distribution of CO for the whole year and by season of birth.
5.11. Distribution of PM_{10} for the whole year and by season of birth.
5.12: Distribution of O_3 for the whole year and by season of birth.
5.13. Targeted maximum likelihood estimates of the difference in FVC had everyone been exposed to above versus below median concentrations of each pollutant during each exposure period for each age group.
5.14. Targeted maximum likelihood estimates of the difference in FEV_1 had everyone been exposed to above versus below median concentrations of each pollutant during each exposure period for each age group.
5.15. Targeted maximum likelihood estimates of the difference in FEF_{25-75} had everyone been exposed to above versus below median concentrations of each pollutant during each exposure period for each age group.
5.16. Targeted maximum likelihood estimates of the difference in FEV_1/FVC had everyone been exposed to above versus below median concentrations of each pollutant during each exposure period for each age group.

5.17. Targeted maximum likelihood estimates of the difference in FEF_{25-75}/FVC had everyone been exposed to above versus below median concentrations of each pollutant during each exposure period for each age group.
5.18. Targeted maximum likelihood estimates of the difference in FEF_{75} had everyone been exposed to above versus below median concentrations of each pollutant during each exposure period for each age group.

Table 5.1. Baseline characteristics of the FACES-LITE population (n=162), Total and by age group.

Demographics at baseline	**N (%)**	
Age at baseline in years, median, IQR (range)	8, 3	(6-11)
Race		
White	67	(41)
Hispanic	62	(38)
African-American	27	(17)
Other	1	(<1)
Male	96	(59)
Family income < $30,000/year	68	(42)
Owns home	101	(62)
Mother's education (completed high school)	138	(85)
Exposures		
Smoker in home	23	(14)
Currently lives within 4 blocks of major roadway	39	(24)
Birth Characteristics		
Low birth weight (<5.5 pounds)	7	(4)
Preterm birth (<37 weeks gestation)	16	(10)
Mother <=18 yr at birth	8	(5)
Mother >=35 yr at birth	26	(16)
First born child	75	(46)
Mother smoked when pregnant	10	(6)
Breastfed any	114	(70)
Born in CA	151	(93)
Born in Fresno or Clovis	127	(78)
Pollutant season of birth		
Winter (Oct-Jan)	47	(29)
Spring(Feb-May)	67	(41)
Summer(June-Sept)	48	(30)
Health history		
Atopy (as measured by skin test)	88	(54)
Asthma diagnosed <= 2 yrs old	60	(37)
Mild Asthma (based on GINA*)	54	(33)
Moderate Asthma (based on GINA*)	77	(48)
Severe Asthma (based on GINA*)	31	(19)
Use of rescue medication	106	(65)
Use of controller medication	66	(41)
Residential History Provided		
Percent of pregnancy, median (range)	100	(11-100)

*Global Initiative for Asthma

Table 5.2. Longitudinal characteristics of the FACES-LITE population, Total and by age group - Mean, standard deviation (SD).

Pulmonary Function Test	**Total (N=947)**		**Ages 6-8 (N=349)**		**Ages 9-10 (N=330)**		**Ages 11-13 (N=268)**	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
FVC (L)	2.29	0.64	1.79	0.38	2.31	0.40	2.92	0.59
FEV_1 (L/s)	1.87	0.54	1.47	0.34	1.87	0.36	2.38	0.51
FEF_{25-75} (L/s)	2.00	0.79	1.62	0.61	1.97	0.66	2.54	0.85
FEV_1/FVC	0.82	0.08	0.82	0.08	0.81	0.08	0.82	0.08
FEF_{25-75}/FVC	0.88	0.28	0.90	0.29	0.86	0.27	0.88	0.27
FEF_{75} (L/s)	0.96	0.48	0.79	0.35	0.95	0.50	1.21	0.47
Anthropometrics	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Height (inches)	54.8	5.4	49.9	3.2	55.2	3.0	60.6	3.5
Weight (pounds)	84.7	33.1	61.0	15.5	86.1	26.4	113.9	33.5

Table 5.3. Correlations between and among 24-hour averages of regulatory pollutant metrics.

	Entire Pregnancy			
Entire Pregnancy	NO_2	CO	PM_{10}	O_3
NO_2	1	0.78	0.34	-0.40
CO		1	0.27	-0.43
PM_{10}			1	0.20
O_3				1
	First Trimester			
Entire Pregnancy	NO_2	CO	PM_{10}	O_3
NO_2	0.71			
CO		0.48		
PM_{10}			0.61	
O_3				0.34
	Second Trimester			
Entire Pregnancy	NO_2	CO	PM_{10}	O_3
NO_2	0.82			
CO		0.76		
PM_{10}			0.76	
O_3				0.69
	Third Trimester			
Entire Pregnancy	NO_2	CO	PM_{10}	O_3
NO_2	0.68			
CO		0.56		
PM_{10}			0.62	
O_3				0.42

Table 5.4. Population median and range of concentrations and numbers of observations above (▲) and below (▼) the median for each pollutant during each interval by age group. (Pollutants were measured in the following units: NO_2 – ppb, PM_{10} – $\mu g/m^3$, CO – ppm, O_3 – ppb.)

Entire Pregnancy									
Pollutant	**Median**	**IQR**	**Range**	**6-8**		**9-10**		**11-13**	
				▲	▼	▲	▼	▲	▼
NO_2	21.03	4.72	8.96-48.56	165	175	164	162	164	92
PM_{10}	45.86	15.56	17.89-94.30	131	218	189	141	208	60
CO	1.33	0.55	0.40-3.58	141	202	172	144	166	84
O_3	47.97	9.86	13.90-77.35	159	186	159	170	163	100
First Trimester									
Pollutant	**Median**	**IQR**	**Range**	**6-8**		**9-10**		**11-13**	
				▲	▼	▲	▼	▲	▼
NO_2	21.00	10.26	7.52-59.80	153	152	159	133	128	102
PM_{10}	42.13	26.83	13.53-114.60	145	166	163	136	141	99
CO	1.15	1.02	0.35-4.00	127	178	160	125	131	91
O_3	49.77	30.88	9.63-84.70	166	139	150	145	122	109
Second Trimester									
Pollutant	**Median**	**IQR**	**Range**	**6-8**		**9-10**		**11-13**	
				▲	▼	▲	▼	▲	▼
NO_2	21.87	10.38	6.35-46.13	164	148	158	144	133	103
PM_{10}	42.33	21.73	14.53-114.30	162	159	167	139	148	97
CO	1.33	1.20	0.35-5.10	163	152	158	134	124	103
O_3	44.33	33.13	14.47-84.70	139	178	155	155	136	104
Third Trimester									
Pollutant	**Median**	**IQR**	**Range**	**6-8**		**9-10**		**11-13**	
				▲	▼	▲	▼	▲	▼
NO_2	20.43	9.90	9.5-54.23	161	178	186	137	152	104
PM_{10}	40.43	18.75	14.70-115.90	138	210	179	148	172	96
CO	1.20	1.13	0.37-5.67	137	205	157	156	140	110
O_3	44.40	33.27	10.80-84.70	151	189	167	159	148	115

Table 5.5. Variables Selected in the g (x) and Q(o) models for each pollutant by pregnancy period and age group

Table 5.5a. Nitrogen Dioxide

Variable	1st Trimester			2nd Trimester			3rd Trimester			Entire Pregnancy		
	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13
Season of birth	o		o	o	xo	x	o			xo	o	
Year of birth	o		o	x	o		xo		xo	o	o	xo
Born in Fresno/Clovis			o					xo	x			
Born in California			o						o			o
First born												
Race	o			o	x		o			o		
Sex			o			o	o		o	x	o	o
Home ownership at baseline	o	o							o	x	o	o
Low income									o	x		o
Maternal age at birth (<18)	x						x	xo				
Maternal age at birth (>35)			x	x			x					
Maternal education							x					
Maternal prenatal smoking			o			o			o			o

Table 5.5b. Carbon Monoxide

Variable	1st Trimester			2nd Trimester			3rd Trimester			Entire Pregnancy		
	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13
Season of birth	xo	x		xo	xo	x	o	x	o	xo	x	o
Year of birth	o	x	o	o		o	xo	xo	o	o	x	xo
Born in Fresno/Clovis	x	x							xo	x		x
Born in California												
First born								x	x			
Race	o			o			xo			xo		
Sex	o	o	o		o	o	xo	o	o		o	xo
Home ownership at baseline			o			o			o	x		xo
Low income				xo			o		o			
Maternal age at birth (<18)	x											
Maternal age at birth (>35)												
Maternal education				xo			o			x		
Maternal prenatal smoking			o			o			o			xo

Table 5.5c. Particle Matter $< 10\mu m$

Variable	1st Trimester			2nd Trimester			3rd Trimester			Entire Pregnancy		
	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13
Season of birth	xo	o		x	xo	x	xo	xo	x	xo	x	
Year of birth	o	o		xo			o	x	xo		xo	x
Born in Fresno/Clovis	x	x	x	x	x	x		x	x	x	xo	x
Born in California								o			o	
First born		o										
Race	o			o			o	o		o		
Sex		o	o		o	o		o	o	o	o	o
Home ownership at baseline		o	xo			o		o	o	x		o
Low income									o			o
Maternal age at birth (<18)	x											
Maternal age at birth (>35)												
Maternal education												
Maternal prenatal smoking			o			o			o			o

Table 5.5d. Ozone

Variable	1st Trimester			2nd Trimester			3rd Trimester			Entire Pregnancy		
	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13
Season of birth	xo	x	x	o			xo	xo		xo	xo	x
Year of birth	o			o			o	o		o	o	x
Born in Fresno/Clovis		x	x							x	x	x
Born in California												
First born												
Race	o			o			o			x		
Sex						o		o	o	o	o	o
Home ownership at baseline						o		o	o			o
Low income				o			o		o			o
Maternal age at birth (<18)												
Maternal age at birth (>35)												
Maternal education				o			o					
Maternal prenatal smoking			o			o			o		o	o

Table 5.6. Results for NO_2: causal inference methods, presented in absolute difference and percent difference, and traditional methods. Deficits in pulmonary function are in **bold**.

Table 5-6a. Causal inference methods: Marginal association between NO_2 exposure during pregnancy and mean pulmonary function by age group. Results compare above versus below median levels of 24-hour averages during pregnancy intervals, based on the targeted maximum likelihood estimation.

Pulmonary Function Test	Entire Pregnancy			1st Trimester			2nd Trimester			3rd Trimester		
	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13
FVC (L)	0.03	**-0.18**	**-0.10**	**-0.07**	**-0.02**	**-0.02**	**-0.01**	**-0.36**	**-0.04**	0.01	**-0.01**	0.14
FEV_1 (L/s)	0.06	**-0.04**	**-0.12**	**-0.01**	**-0.07**	**-0.12**	**-0.14**	**-0.25**	**-0.09**	**-0.00**	0.02	0.10
FEF_{25-75}(L/s)	0.13	**-0.03**	**-0.44**	**-0.02**	**-0.17**	**-0.40**	**-0.55**	**-0.26**	**-0.28**	**-0.04**	0.06	**-0.10**
FEV_1/FVC	0.01	0.01	**-0.01**	0.00	**-0.02†**	**-0.06***	**-0.07†**	0.00	**-0.02**	**-0.01**	0.01	**-0.02**
FEF_{25-75}/FVC	0.04	**-0.01**	**-0.04**	**-0.00**	**-0.07**	**-0.23***	**-0.31***	0.01	**-0.08**	**-0.04**	0.03	**-0.10**
FEF_{75}(L/s)	0.05	**-0.09**	**-0.33**	**-0.01**	**-0.15**	**-0.27**	**-0.35†**	**-0.16**	**-0.22**	**-0.04**	0.06	0.04

*Significant at 95% confidence level
†Significant at 90% confidence level

Table 5.6b. Same results as above (NO_2) presented as *percent change* in pulmonary function by age group.

Pulmonary Function Test	Entire Pregnancy			1st Trimester			2nd Trimester			3rd Trimester		
	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13
FVC (L)	0.02	**-0.08**	**-0.03**	**-0.04**	**-0.01**	**-0.01**	**-0.01**	**-0.16**	**-0.01**	0.01	0.00	0.05
FEV_1 (L/s)	0.04	**-0.02**	**-0.05**	**-0.01**	**-0.04**	**-0.05**	**-0.10**	**-0.13**	**-0.04**	0.00	0.01	0.04
FEF_{25-75}(L/s)	0.08	**-0.02**	**-0.17**	**-0.01**	**-0.09**	**-0.16**	**-0.34**	**-0.13**	**-0.11**	**-0.02**	0.03	**-0.04**
FEV_1/FVC	0.01	0.01	**-0.01**	0.00	**-0.02†**	**-0.07***	**-0.09†**	0.00	**-0.02**	**-0.01**	0.01	**-0.02**
FEF_{25-75}/FVC	0.04	**-0.01**	**-0.05**	0.00	**-0.08**	**-0.26***	**-0.34***	0.01	**-0.09**	**-0.04**	0.03	**-0.11**
FEF_{75}(L/s)	0.06	**-0.09**	**-0.27**	**-0.01**	**-0.16**	**-0.22**	**-0.44†**	**-0.17**	**-0.18**	**-0.05**	0.06	0.03

Table 5.6c. Traditional method: Conditional association between NO_2 exposure during pregnancy and mean pulmonary function by age group (conditional on all variables in Table 5.5a).

Pulmonary Function Test	Entire Pregnancy			1st Trimester			2nd Trimester			3rd Trimester		
	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13
FVC (L)	0.05	-0.10*	-0.27*	-0.02	0.01	0.34*	-0.04	-0.23*	-0.03	0.02	-0.03	-0.19
FEV_1 (L/s)	0.04	-0.08†	-0.32*	-0.02	-0.10	-0.39*	-0.14	-0.19*	-0.03	0.00	0.03	-0.21
FEF_{25-75}(L/s)	0.03	-0.08	-0.54*	-0.06	-0.35	-0.07*	-0.51*	-0.23†	-0.10	-0.08	0.15	-0.35
FEV_1/FVC	0.00	0.00	-0.03*	-0.01	-0.05	-0.26*	-0.06*	-0.01	-0.01	-0.01	0.02	-0.01
FEF_{25-75}/FVC	-0.02	0.00	-0.10*	-0.04	-0.16	-0.21*	-0.29*	-0.03	-0.04	-0.07†	0.07	-0.04
FEF_{75}(L/s)	-0.01	-0.13†	-0.34*	-0.05	-0.22	0.00	-0.29*	-0.19†	-0.06	-0.07	0.07	-0.23

*P < 0.05
†P < 0.10

Table 5.7. Results for CO: causal inference methods, presented in absolute difference and percent difference, and traditional methods. Deficits in pulmonary function are in **bold**.

Table 5.7a. Causal inference methods: Marginal association between CO exposure during pregnancy and mean pulmonary function by age group. Results compare above versus below median levels of 24-hour averages during pregnancy intervals, based on the targeted maximum likelihood estimation

Pulmonary Function Test	Entire Pregnancy			1st Trimester			2nd Trimester			3rd Trimester		
	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13
FVC (L)	0.00	**-0.01**	**-0.05**	0.04	0.15	0.01	0.04	**-0.10**	**-0.05**	0.06	0.03	0.08
FEV_1 (L/s)	**-0.01**	0.01	**-0.09**	0.03	0.11	**-0.01**	0.00	**-0.12**	**-0.12**	0.05	0.02	0.07
FEF_{25-75}(L/s)	**-0.07**	0.06	**-0.09**	0.02	0.01	**-0.03**	**-0.03**	**-0.19**	**-0.16**	**-0.01**	0.03	**-0.01**
FEV_1/FVC	**-0.02**†	0.02*	**-0.01**	**-0.04***	**-0.01**	**-0.01**	**-0.02***	**-0.04***	**-0.06**	**-0.01**	0.00	0.04
FEF_{25-75}/FVC	**-0.09**	0.08	0.00	**-0.11**	**-0.13**	**-0.03**	**-0.08**	**-0.09**	**-0.08**	**-0.08**	0.01	0.06
FEF_{75}(L/s)	**-0.09**	0.04	**-0.08**	0.08	0.00	0.06	**-0.03**	**-0.20**	**-0.27**	**-0.09**	0.04	0.02

*Significant at 95% confidence level, calculated from the influence curve
†Significant at 90% confidence level, calculated from the influence curve

Table 5.7b. Same results as above (CO) presented as *percent change* in pulmonary function by age group.

Pulmonary Function Test	Entire Pregnancy			1st Trimester			2nd Trimester			3rd Trimester		
	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13
FVC (L)	0.00	**-0.02**	**-0.15**	0.08	0.35	0.02	0.06	**-0.24**	**-0.16**	0.11	0.06	0.24
FEV_1 (L/s)	**-0.01**	0.02	**-0.22**	0.05	0.21	**-0.03**	0.00	**-0.23**	**-0.29**	0.08	0.04	0.17
FEF_{25-75}(L/s)	**-0.11**	0.12	**-0.22**	0.03	0.02	**-0.08**	**-0.05**	**-0.37**	**-0.41**	**-0.01**	0.05	**-0.03**
FEV_1/FVC	**-0.02**†	0.02*	**-0.01**	**-0.03***	**-0.01**	**-0.01**	**-0.02***	**-0.03***	**-0.05**	**-0.01**	-0.00	0.03
FEF_{25-75}/FVC	**-0.08**	0.07	0.00	**-0.10**	**-0.11**	**-0.03**	**-0.07**	**-0.08**	**-0.07**	**-0.07**	0.01	0.05
FEF_{75}(L/s)	**-0.07**	0.04	**-0.1**	0.06	0.00	0.07	**-0.02**	**-0.19**	**-0.33**	**-0.07**	0.04	0.03

Table 5.7c. Traditional method: Conditional association between CO exposure during pregnancy and mean pulmonary function by age group. (*conditional on all variables in Table 5.5b)

Pulmonary Function Test	Entire Pregnancy			1st Trimester			2nd Trimester			3rd Trimester		
	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13
FVC (L)	0.00	**-0.08**	**-0.24***	0.00	0.20*	0.44*	0.07	**-0.11**	**-0.01**	0.09	0.01	**-0.13**
FEV_1 (L/s)	**-0.03**	-0.03	**-0.16†**	0.02	0.14*	0.31*	**-0.01**	**-0.17***	0.00	0.02	0.03	**-0.06**
FEF_{25-75}(L/s)	**-0.15***	0.07	**-0.07**	0.11	0.02	0.24	**-0.21***	**-0.35***	**-0.01**	**-0.14**	0.09	0.01
FEV_1/FVC	**-0.02***	0.02	0.02	0.02†	**-0.01**	0.02	**-0.04***	**-0.04***	0.00	**-0.03***	0.01	0.02
FEF_{25-75}/FVC	**-0.10***	0.06	0.07†	0.07	**-0.05**	-0.06	**-0.14***	**-0.11***	**-0.02**	**-0.15***	0.04	0.05
FEF_{75}(L/s)	**-0.10***	0.02	**-0.06**	0.00	**-0.02**	0.21*	**-0.11***	**-0.21†**	**-0.01**	**-0.15***	0.08	0.06

*P < 0.05

†P < 0.10

Table 5.8. Results for PM_{10}: causal inference methods, presented in absolute difference and percent difference, and traditional methods. Deficits in pulmonary function are in **bold**.

Table 5.8a. Causal inference methods: Marginal association between PM_{10} exposure during pregnancy and mean pulmonary function by age group. Results compare above versus below median levels of 24-hour averages during pregnancy intervals, based on the targeted maximum likelihood estimation.

Pulmonary Function Test	Entire Pregnancy			1st Trimester			2nd Trimester			3rd Trimester		
	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13
FVC (L)	**-0.05**	**-0.14**	0.03	**-0.04**	**-0.05**	**-0.15**	0.03	**-0.05**	0.05	**-0.08**	**-0.06**	0.25
FEV_1 (L/s)	**-0.03**	**-0.05**	0.29	**-0.05**	**-0.03**	**-0.14**	**-0.06**	**-0.01**	0.06	**-0.01**	0.03	0.21
FEF_{25-75}(L/s)	**-0.08**	0.12	0.16	0.07	0.02	**-0.15**	**-0.16**	0.00	0.35	0.14	0.21	0.34
FEV_1/FVC	0.00	0.02	0.03*	0.01	0.01	**-0.01**	**-0.02***	0.02	0.04	0.02*	0.03*	0.03
FEF_{25-75}/FVC	**-0.02**	0.11	0.03	0.07	0.03	0.02	**-0.09**	0.02	0.10	0.09	0.10	**-0.02**
FEF_{75}(L/s)	**-0.01**	0.05	0.18	0.09	0.03	**-0.15**	**-0.11**	0.07	0.01	0.06	0.15	**-0.08**

*Significant at 95% confidence level, calculated from the influence curve
†Significant at 90% confidence level, calculated from the influence curve

Table 5.8b. Same results as above (PM_{10}) presented as *percent change* in pulmonary function by age group.

Pulmonary Function Test	Entire Pregnancy			1st Trimester			2nd Trimester			3rd Trimester		
	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13
FVC (L)	**-0.03**	**-0.06**	0.01	**-0.02**	**-0.02**	**-0.05**	0.02	**-0.02**	0.02	**-0.04**	**-0.03**	0.09
FEV_1 (L/s)	**-0.02**	**-0.03**	0.12	**-0.03**	**-0.02**	**-0.06**	**-0.04**	**-0.01**	0.03	**-0.01**	0.02	0.09
FEF_{25-75}(L/s)	**-0.05**	0.06	0.06	0.04	0.01	**-0.06**	**-0.1**	0.00	0.14	0.09	0.11	0.13
FEV_1/FVC	0.00	0.02	0.04*	0.01	0.01	**-0.01**	**-0.02***	0.02	0.05	0.02*	0.04*	0.04
FEF_{25-75}/FVC	**-0.02**	0.13	0.03	0.08	0.03	0.02	**-0.1**	0.02	0.11	0.10	0.12	**-0.02**
FEF_{75}(L/s)	**-0.01**	0.05	0.15	0.11	0.03	**-0.12**	**-0.14**	0.07	0.01	0.08	0.16	**-0.07**

Table 5.8c. Traditional method: Conditional association between PM_{10} exposure during pregnancy and mean pulmonary function by age group (conditional on all variables in Table 5.5c).

Pulmonary Function Test	Entire Pregnancy			1st Trimester			2nd Trimester			3rd Trimester		
	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13
FVC (L)	-0.01	-0.13*	0.00	-0.03	-0.06	0.13	0.06	-0.07	0.11	-0.09†	-0.11	-0.18†
FEV_1 (L/s)	-0.02	-0.07	0.14	-0.02	-0.06	0.08	0.01	-0.03	0.21*	-0.03	0.01	-0.09
FEF_{25-75}(L/s)	-0.10	0.01	0.42*	0.02	-0.05	0.04	-0.13	0.03	0.40*	0.05	0.23*	-0.01
FEV_1/FVC	-0.01	0.01	0.05*	0.00	0.00	-0.01	-0.02*	0.01	0.04*	0.02	0.04*	0.03*
FEF_{25-75}/FVC	-0.06	0.05	0.15*	0.02	0.01	-0.02	-0.12*	0.03	0.10*	0.05	0.13*	0.07†
FEF_{75}(L/s)	-0.04	0.01	0.23*	0.06	-0.03	0.00	-0.09†	0.14	0.28*	0.02	0.19*	-0.06

*P < 0.05
†P < 0.10

Table 5.9. Results for O_3: causal inference methods, presented in absolute difference and percent difference, and traditional methods. Deficits in pulmonary function are in **bold**.

Table 5-9a. Causal inference methods: Marginal association between O_3 exposure during pregnancy and mean pulmonary function by age group. Results compare above versus below median levels of 24-hour averages during pregnancy intervals, based on the targeted maximum likelihood estimation.

Pulmonary Function Test	Entire Pregnancy			1st Trimester			2nd Trimester			3rd Trimester		
	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13
FVC (L)	**-0.01**	0.07	0.02	0.01	**-0.03**	0.09	0.04	0.08	**-0.05**	**-0.10**	**-0.09**	**-0.10**
FEV_1 (L/s)	**-0.06**	0.04	**-0.01**	**-0.07**	**-0.02**	0.15	**-0.04**	0.11	0.02	**-0.06**	0.08	**-0.02**
FEF_{25-75}(L/s)	**-0.21**	0.03	0.08	**-0.28**	**-0.03**	0.28	**-0.03**	0.21	0.14	**-0.01**	0.08	0.16
FEV_1/FVC	**-0.02***	0.00	**-0.00**	**-0.04***	0.00	0.03	0.00	0.02	0.03	0.02*	**-0.00**	0.01
FEF_{25-75}/FVC	**-0.08**	**-0.01**	**-0.01**	**-0.14**	0.01	0.12	**-0.01**	0.06	0.10	0.09	**-0.01**	**-0.03**
FEF_{75}(L/s)	**-0.08**	0.07	**-0.019**	**-0.16**	**-0.02**	0.17	**-0.03**	0.13	**-0.05**	0.05	0.08	0.10

*Significant at 95% confidence level, calculated from the influence curve
†Significant at 90% confidence level, calculated from the influence curve

Table 5.9b. Same results as above (O_3) presented as *percent change* in pulmonary function by age group.

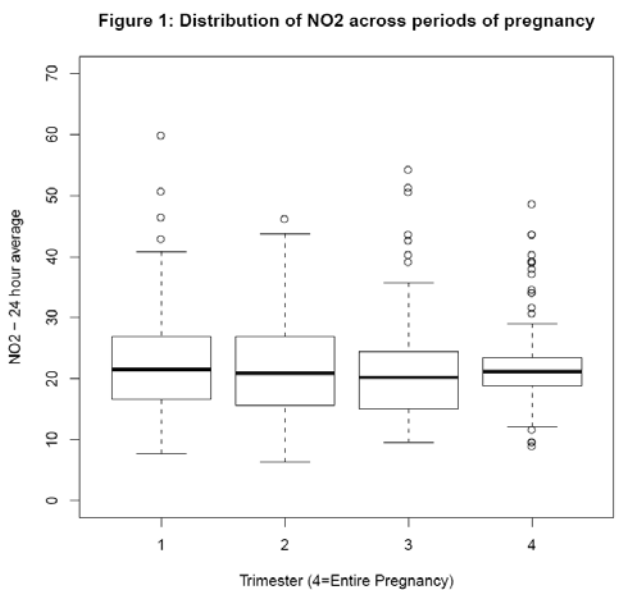
Pulmonary Function Test	Entire Pregnancy			1st Trimester			2nd Trimester			3rd Trimester		
	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13
FVC (L)	**-0.01**	0.03	0.01	0.01	**-0.01**	0.03	0.02	0.03	**-0.02**	**-0.06**	**-0.04**	**-0.03**
FEV_1 (L/s)	**-0.04**	0.02	0.00	**-0.05**	**-0.01**	0.06	**-0.03**	0.06	0.01	**-0.04**	0.04	**-0.01**
FEF_{25-75}(L/s)	**-0.13**	0.02	0.03	**-0.17**	**-0.02**	0.11	**-0.02**	0.11	0.06	**-0.01**	0.04	0.06
FEV_1/FVC	**-0.02***	0.00	0.00	**-0.05***	0.00	0.04	0.00	0.02	0.04	0.02*	0.00	0.01
FEF_{25-75}/FVC	**-0.09**	**-0.01**	**-0.01**	**-0.16**	0.01	0.14	**-0.01**	0.07	0.11	0.10	**-0.01**	**-0.03**
FEF_{75}(L/s)	**-0.10**	0.07	**-0.02**	**-0.20**	**-0.02**	0.14	**-0.04**	0.14	-0.04	0.06	0.08	0.08

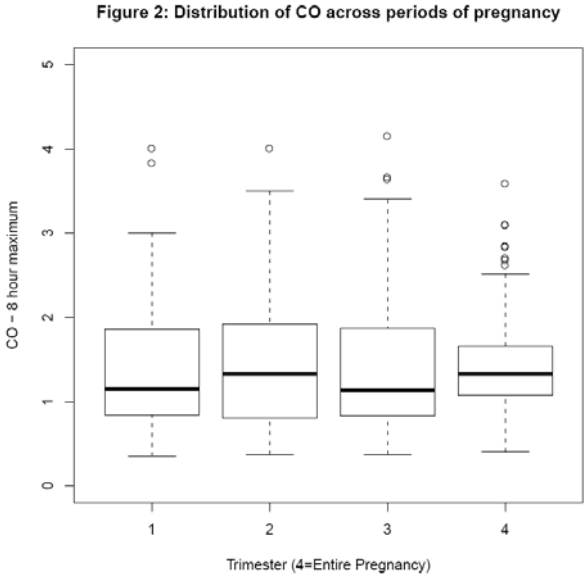
Table 5.9c. Traditional method: Conditional association between O_3 exposure during pregnancy and mean pulmonary function by age group (conditional on all variables in Table 5.5d).

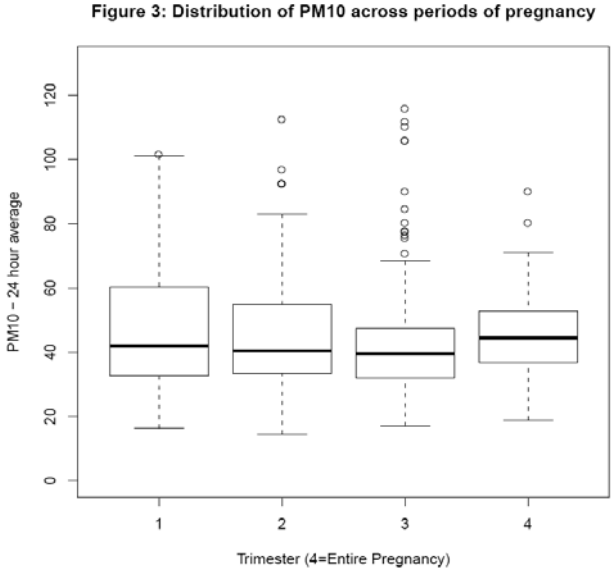
Pulmonary Function Test	Entire Pregnancy			1st Trimester			2nd Trimester			3rd Trimester		
	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13	6-8	9-10	11-13
FVC (L)	**-0.05**	0.00	0.13	-0.05	0.04	0.13	**-0.01**	0.02	**-0.28***	**-0.11***	0.16*	0.26*
FEV_1 (L/s)	**-0.09***	0.01	0.12	-0.09	0.01	0.17†	0.00	0.06	**-0.14**	**-0.04**	0.11*	0.26*
FEF_{25-75}(L/s)	**-0.24***	0.03	0.11	-0.19†	**-0.02**	0.29*	**-0.01**	0.11	0.05	0.07	0.04	0.34*
FEV_1/FVC	**-0.03***	0.00	0.00	-0.02	**-0.01**	0.02	0.00	0.02	0.04*	0.03	**-0.01**	0.01
FEF_{25-75}/FVC	**-0.12***	0.00	0.00	-0.06	**-0.01**	0.05	**-0.01**	0.03	0.12	0.11*	**-0.05**	0.02
FEF_{75}(L/s)	**-0.12***	0.09	0.06	-0.14*	0.05	0.14	**-0.04**	0.09	**-0.06**	0.09†	0.07	0.22*

*P < 0.05
†P < 0.10

Figure 5.1-5.4: Distribution of each pollutant during each trimester (1, 2, 3) and the entire pregnancy (4).







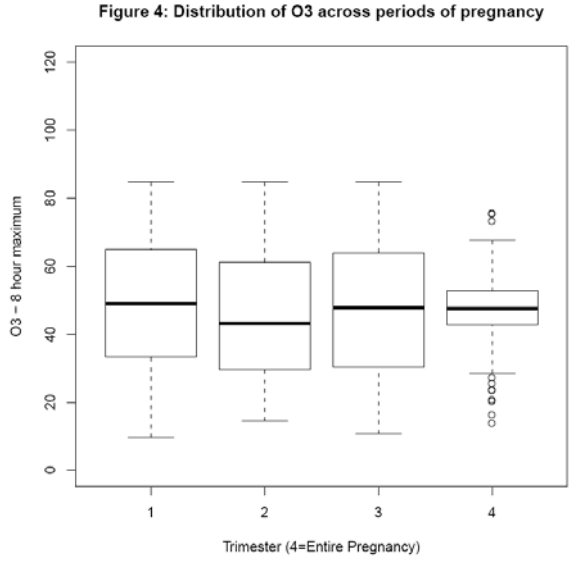
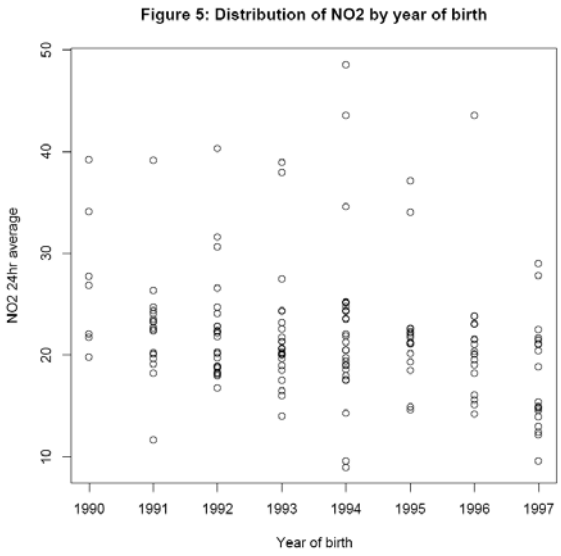
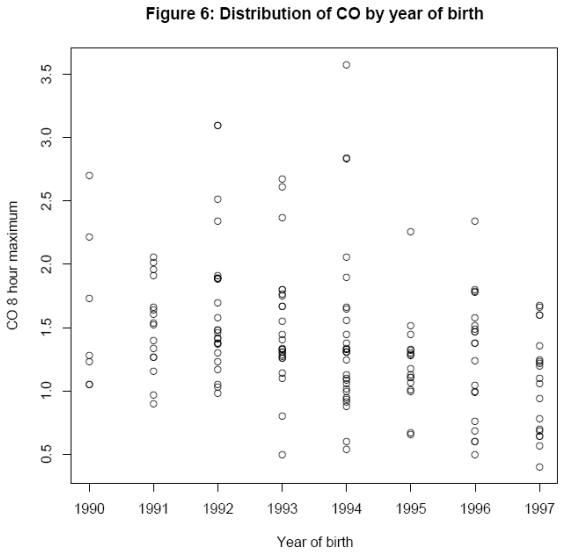
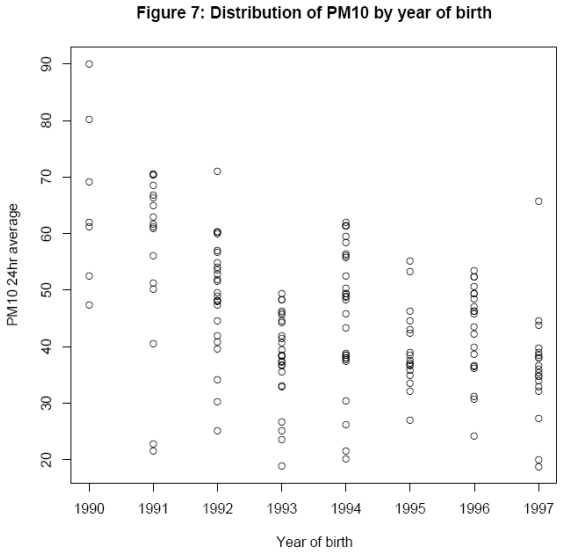


Figure 5.5-5.8: Distribution of each pollutant during the entire pregnancy by year of birth.







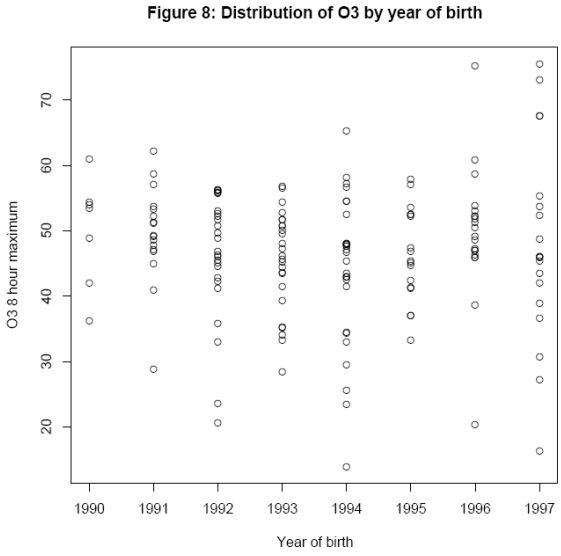
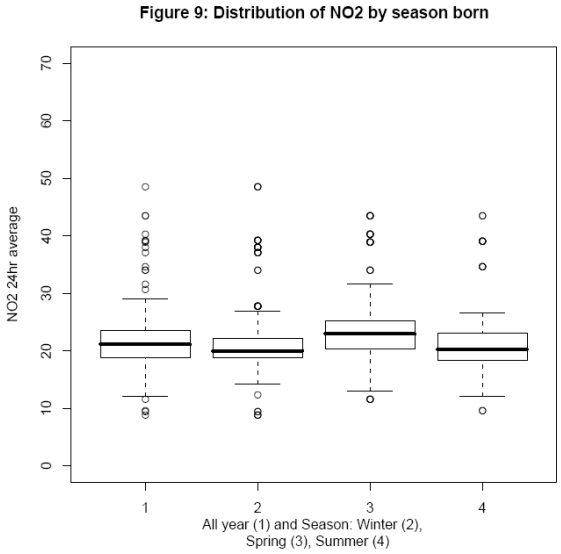
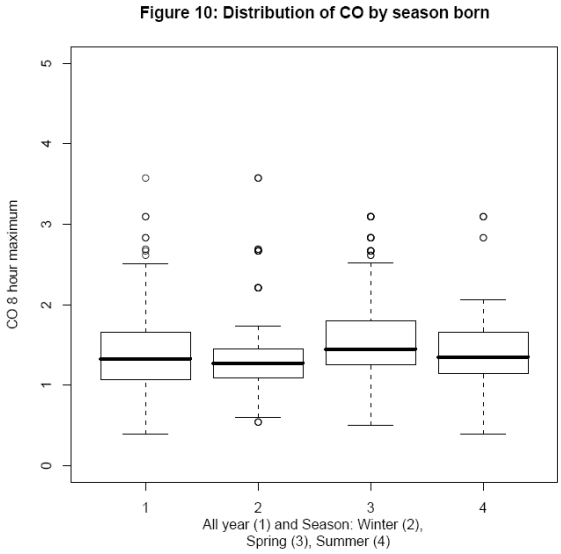
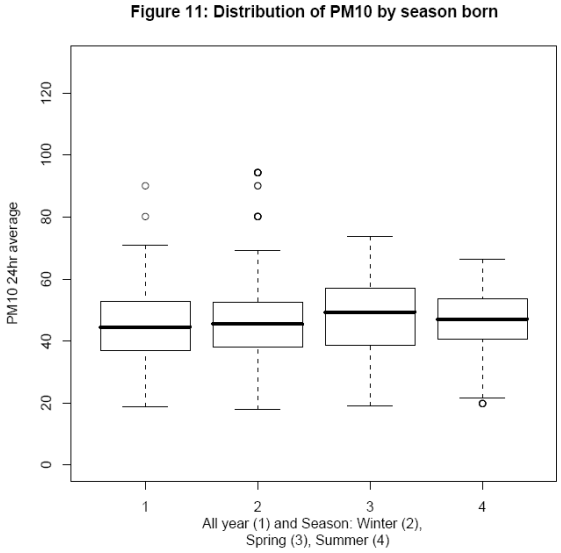


Figure 5.9-5.12: Distribution of each pollutant for the whole year and by season of birth.







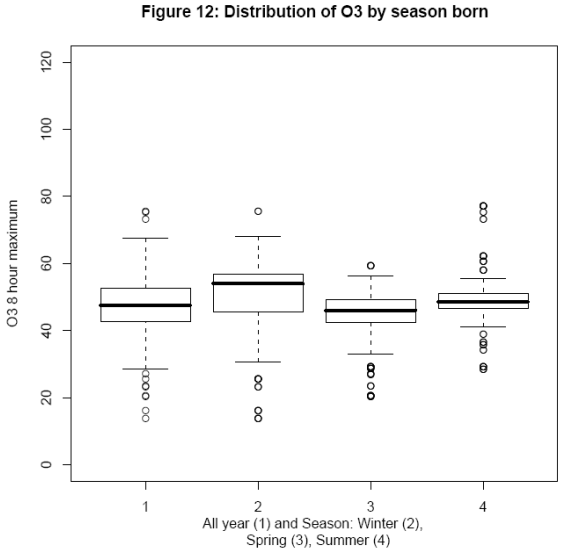


Figure 5.13. Targeted maximum likelihood estimates of the difference in FVC had everyone been exposed to above versus below median concentrations of each pollutant during each exposure period for each age group.

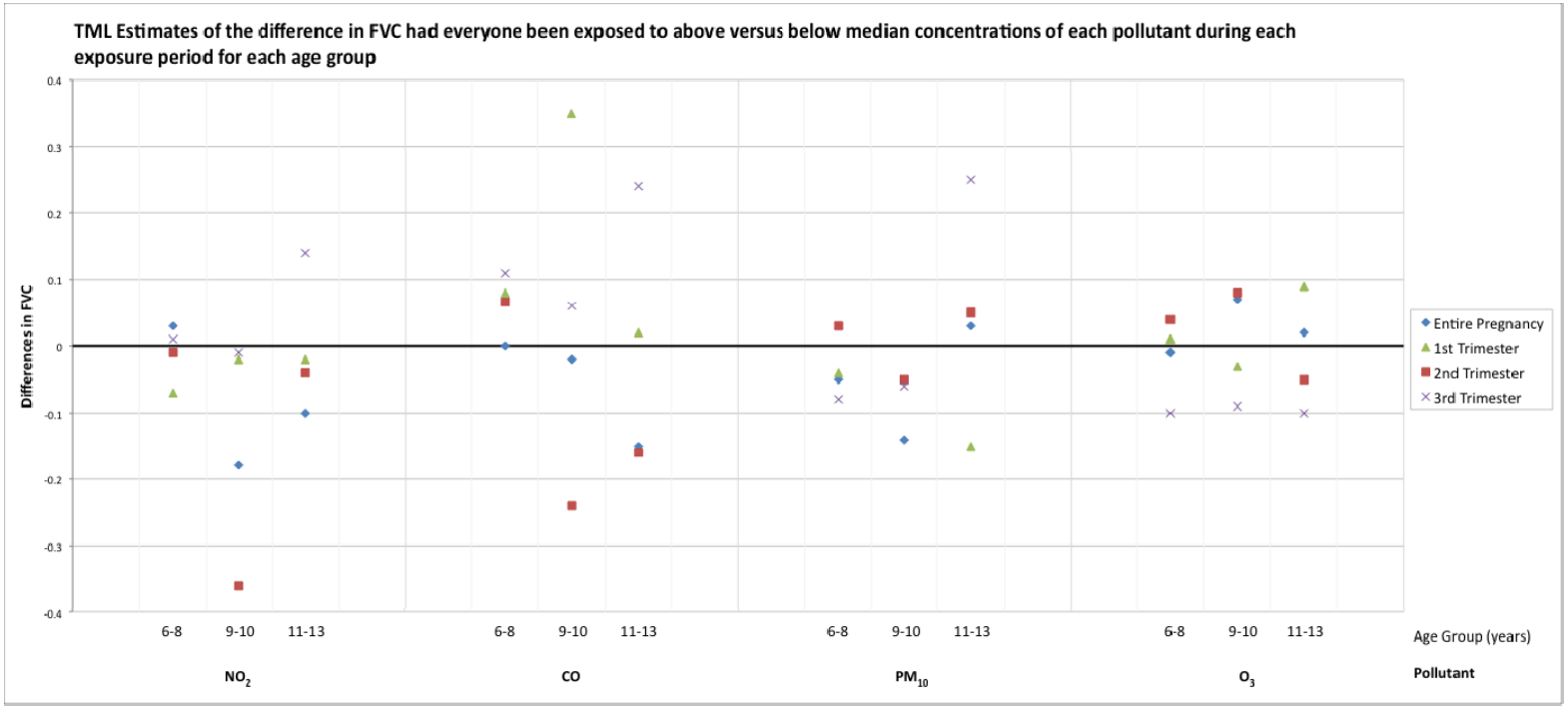


Figure 5.14. Targeted maximum likelihood estimates of the difference in FEV_1 had everyone been exposed to above versus below median concentrations of each pollutant during each exposure period for each age group.

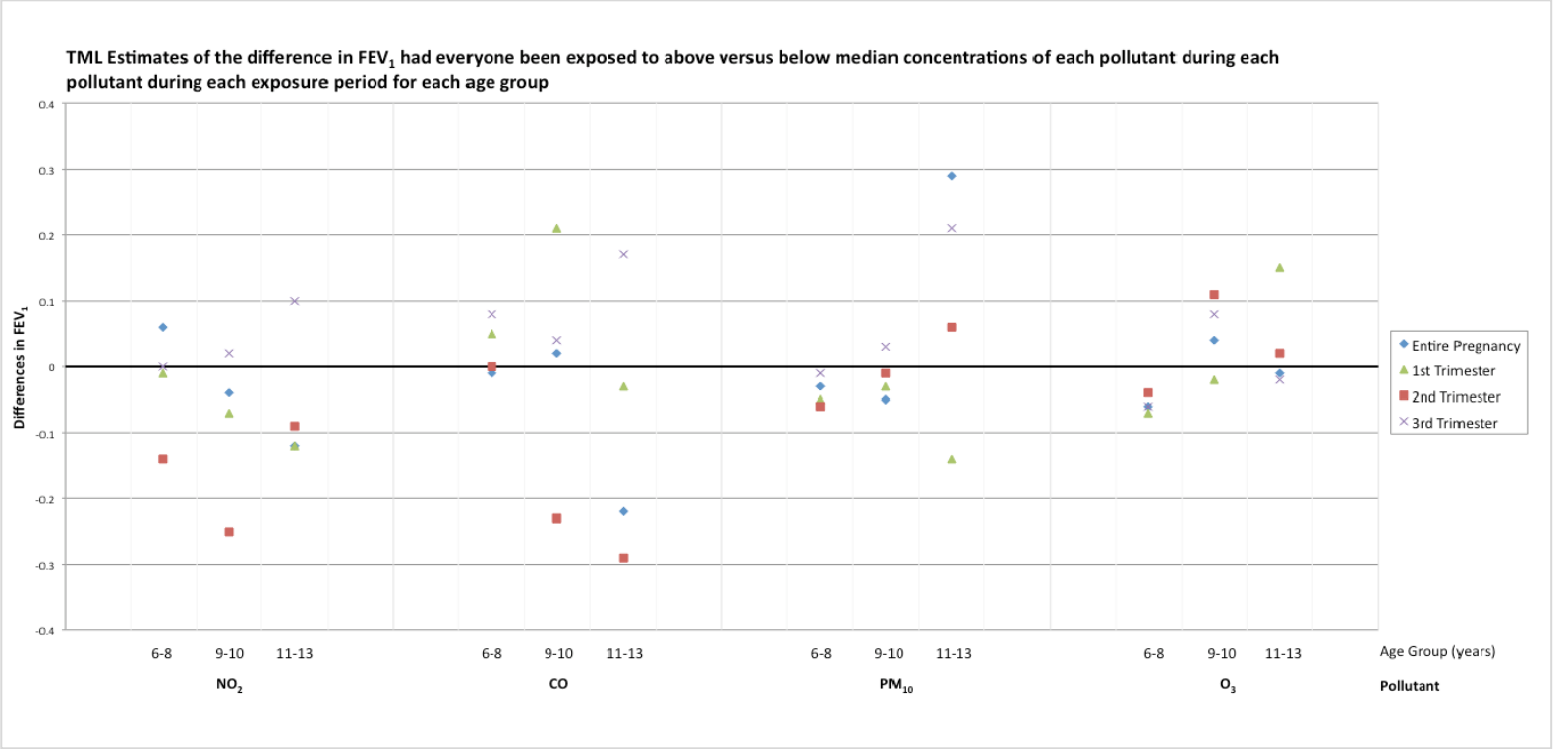


Figure 5.15. Targeted maximum likelihood estimates of the difference in FEF_{25-75} had everyone been exposed to above versus below median concentrations of each pollutant during each exposure period for each age group.

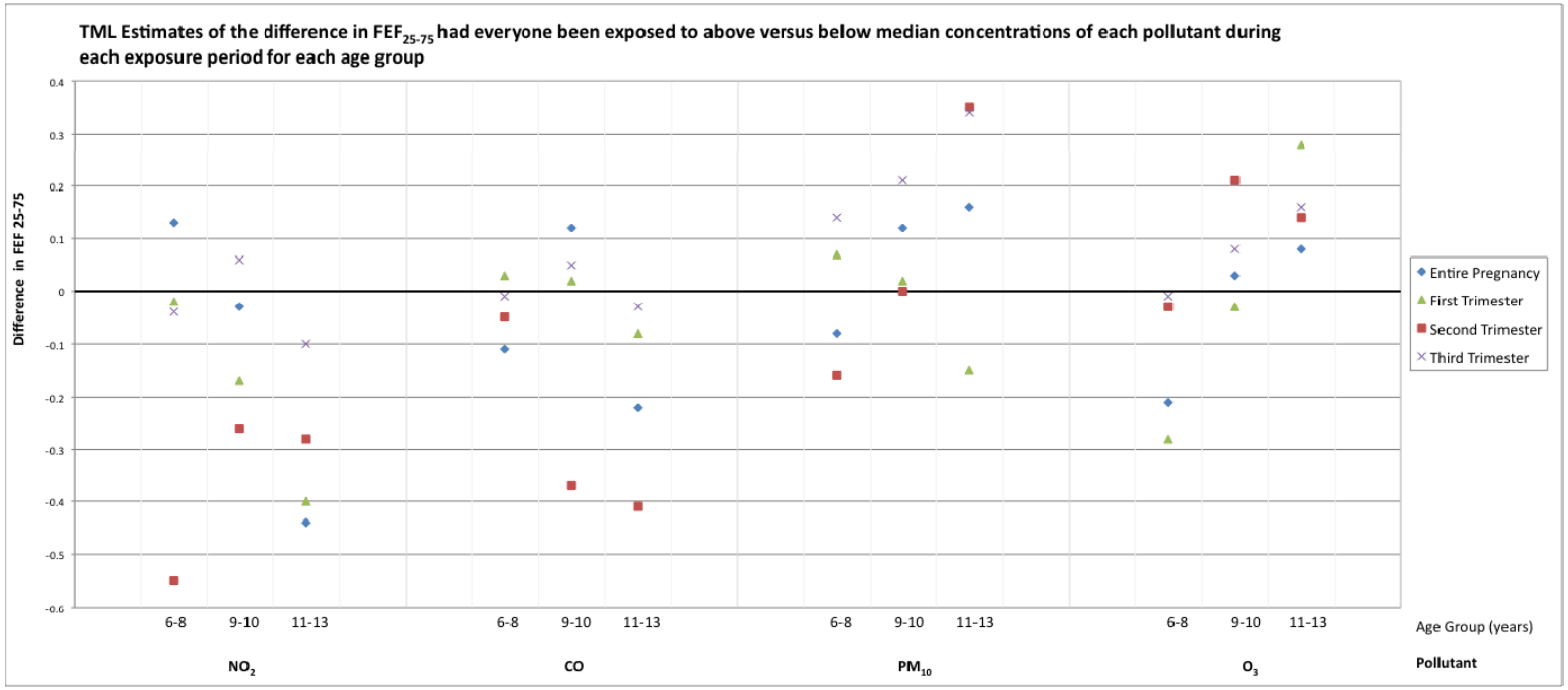


Figure 5.16. Targeted maximum likelihood estimates of the difference in FEV_1/FVC had everyone been exposed to above versus below median concentrations of each pollutant during each exposure period for each age group.

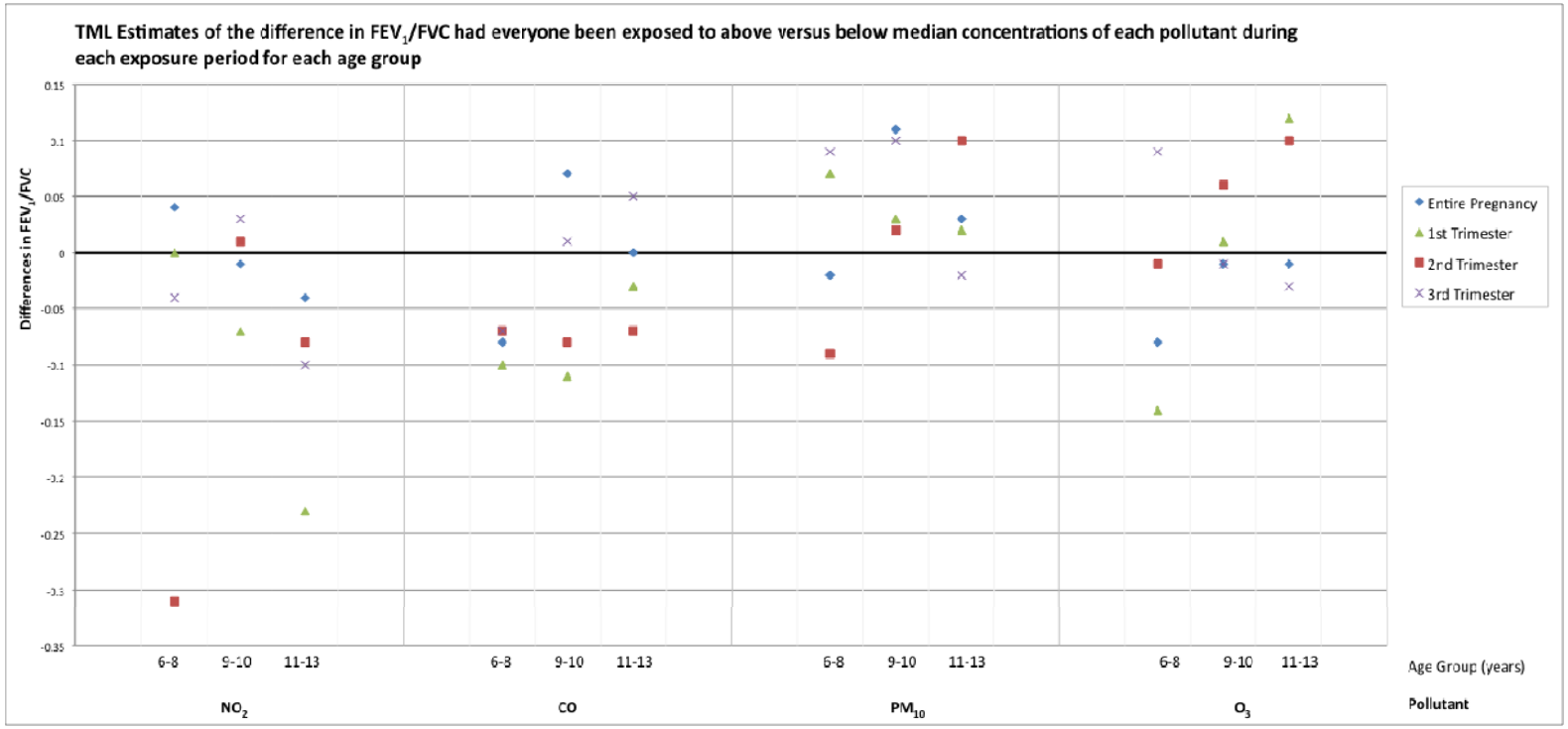


Figure 5.17. Targeted maximum likelihood estimates of the difference in FEF_{25-75}/FVC had everyone been exposed to above versus below median concentrations of each pollutant during each exposure period for each age group.

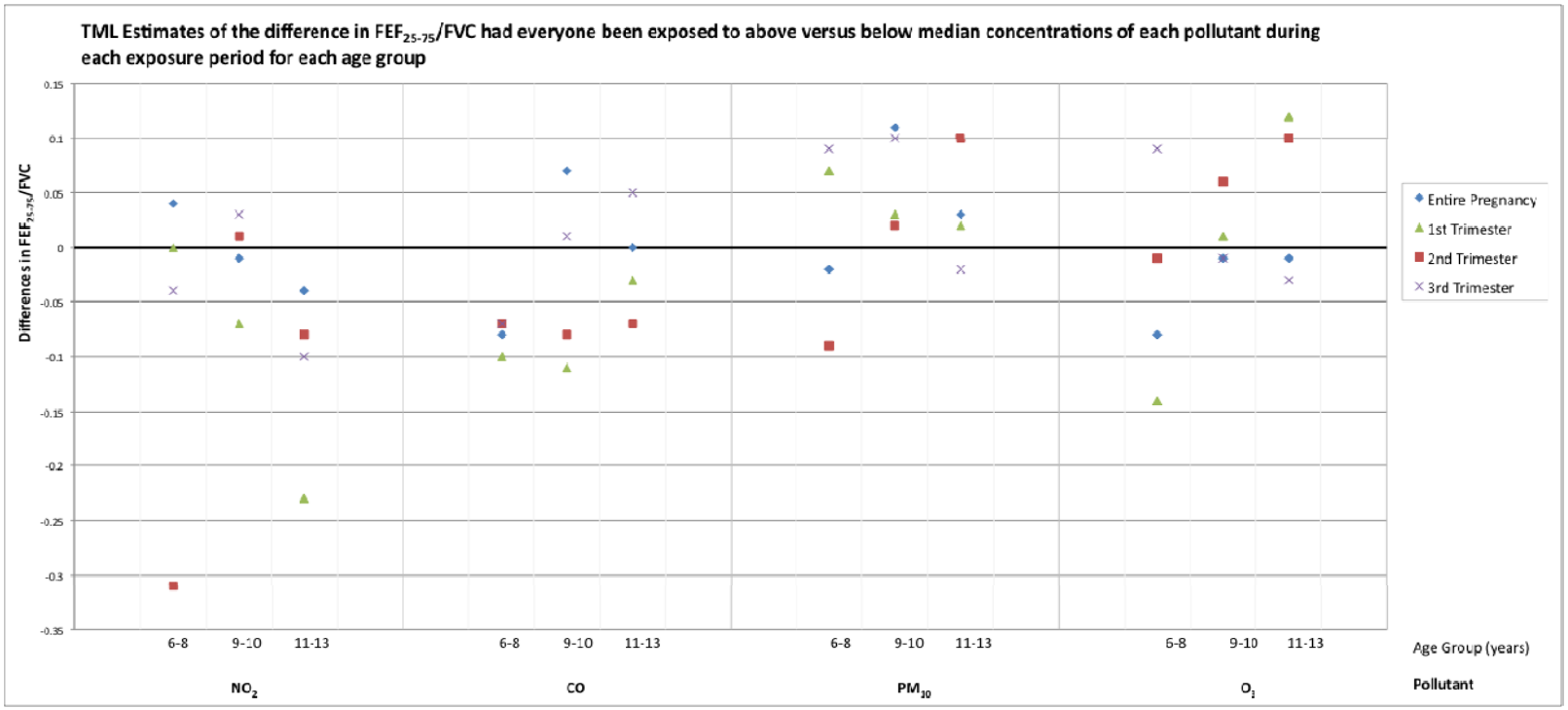
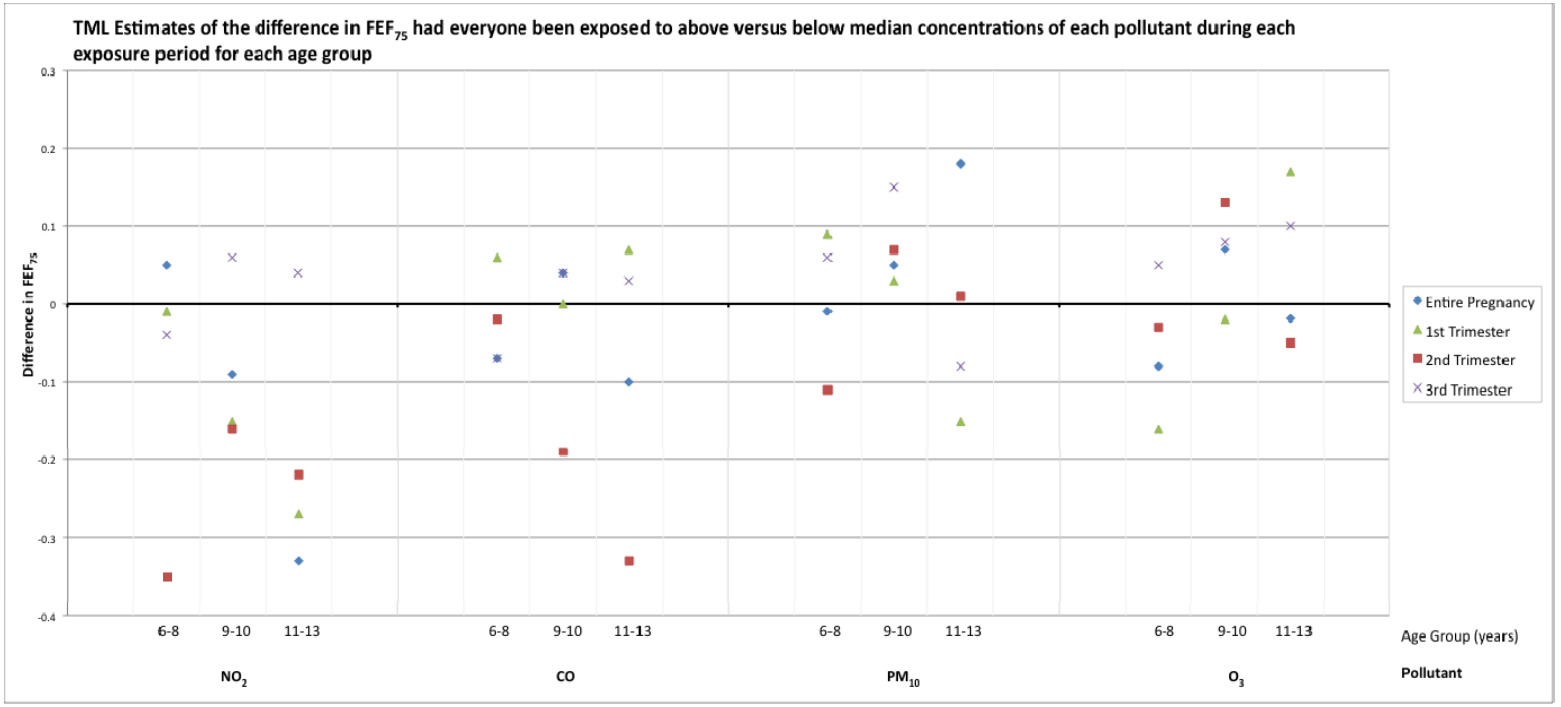


Figure 5.18. Targeted maximum likelihood estimates of the difference in FEF_{75} had everyone been exposed to above versus below median concentrations of each pollutant during each exposure period for each age group.



Chapter 6: The Association Between Exposure to Traffic Density During Pregnancy and Term Low Birth Weight

6.1 Background

As discussed in Chapter 1, ambient air pollution is recognized as an important health problem in the United States and around the world. Motor vehicles are a major source of ambient air pollution in the U.S. Although progress has been made in reducing emissions from individual vehicles, the number of vehicles and miles traveled in the U.S. has grown substantially in the past 15 years (HEI 2010). The expansion of metropolitan areas (urban sprawl) has increased the travel distances between residential and commercial sites and the automobile is the primary means of travel. This growth and change in land use has increased the relative contribution of traffic to the urban pollution mixture.

Analyses of epidemiologic data suggest that prenatal exposure to ambient air pollutants may be associated with a variety of adverse birth outcomes. Traffic-related air pollution has been associated with intrauterine mortality (Pereira, Loomis et al. 1998), low birth weight (Ritz and Yu 1999; Wang, Ding et al. 1997), preterm birth (PTB) (Ritz, Yu et al. 2000), small for gestational age (Dejmek, Solansky et al. 2000; Liu, Krewski et al. 2003), and neonatal mortality (Loomis, Castillejos et al. 1999), as well as postnatal mortality (Woodruff, Grillo et al. 1997; Bobak 2000). However, there is not sufficient evidence to conclude the relation between traffic-related air pollution and birth outcomes is causal. Chapter 3 has a more thorough review of the literature on prenatal effects of air pollution on birth outcomes.

As mentioned in Chapter 3.1, low birth weight (LBW) is classified as weight less than 2500 g (5 lbs. 8 oz.). In 2006, 8.3% of infants in the United States were LBW (Martin, Brown et al. 2009). This prevalence has increased from 6.7% in 1984 (Hamilton, Minino et al. 2007). African-Americans have the highest prevalence of low birth weight (11.9%) compared to other races (Martin, Brown et al. 2009). The prevalence of LBW infants continues to increase in the U.S. (Behrman 2007; Stillerman, Mattison et al. 2008) despite declines in the 1970s and early 1980s. Although recent increases in multiple births have influenced the rise, the prevalence of LBW also is increasing among singleton births (Stillerman, Mattison et al. 2008). Adverse birth outcomes have a financial and emotional burden on families both in the short and long term.

Birth weight and gestational age are important predictors of infant survival and morbidity (Glinianaia, Rankin et al. 2004.). The majority of infant deaths are among infants born with LBW (Wang, Ding et al. 1997). It has been estimated that the lifelong cost associated with a low birth weight infant is $436,000 (Wong, Gohlke et al. 2004). Low birth weight and preterm birth are also important indicators of future health and may play a role in the development of chronic diseases throughout life (Maisonet, Correa et al. 2004). Existing evidence links impaired prenatal growth with serious adult illnesses such as non-insulin-dependent diabetes, hypertension, and coronary heart disease (Barker, Gluckman et al. 1993; Law 1995; Barker 1997).

In Chapter 3.5, I discuss that pregnancy is a critical period of exposure because of the growth and development of the fetus. Pregnancy may constitute a period of human development particularly susceptible to toxins contained in air pollution because of high cell

proliferation, organ development and the changing capabilities of fetal metabolism (Selevan, Kimmel et al. 2000; Ritz and Wilhelm 2008). There are a number of potential mechanisms by which traffic-related air pollution exposure during pregnancy may affect LBW and PTB. Air pollution may affect maternal respiratory or general health and, in turn, impair uteroplacental and umbilical blood flow, transplacental glucose and oxygen transport, all known as major determinants of fetal growth (Ritz and Wilhelm 2008). Additionally, the pollutants to which the fetus is exposed to may cause:

- oxidative stress, which can affect the embryo in the earliest phase of growth,
- inflammation of pulmonary and placental cells, which can induce DNA damage, and
- changes in blood coagulation or hemodynamic responses (Kannan, Misra et al. 2006; Kannan, Misra et al. 2007; Ritz and Wilhelm 2008).

In the past decade, a large number of studies have examined the effects of traffic-related air pollution on birth outcomes such as low birth weight and preterm birth (Dejmek, Selevan et al. 1999; Ritz and Yu 1999; Bobak 2000; Maisonet, Bush et al. 2001; Maroziene and Grazuleviciene 2002; Ha, Lee et al. 2003; Liu, Krewski et al. 2003; Salam, Millstein et al. 2005; Wilhelm and Ritz 2005; Dugandzic, Dodds et al. 2006; Wang, Ding et al. 1997). I focus on one study in particular because of its location in California, which is the focus of this dissertation. Ritz *et al.* examined the associations between ambient air pollution and outcomes among births in Los Angeles, CA from 1989-1993 (Ritz and Yu 1999; Ritz, Yu et al. 2000). Birth certificates were used in a retrospective cohort study of LBW and PTB. Exposures to CO, NO_2 and PM_{10}, were estimated based on data from monitoring stations closest to the mothers' residences. For the birth weight analysis, Ritz *et al.* excluded multiple births and those shorter than 37 or longer than 44 weeks gestation and birth weights greater than 5500 grams and less than 1000 grams. Women with hypertension, diabetes, and uterine bleeding during pregnancy were also excluded from the analysis. Based on logistic regression analysis, they found that those exposed to high levels of CO ($>95^{th}$ percentile; 5.5 ppm) during the third trimester had a relative risk of 1.22 [95% CI: 1.03-1.44] for LBW compared with those exposed to low levels of CO ($<50^{th}$ percentile) (Ritz and Yu 1999). Although this study provided a large sample from an area that is highly polluted by California and national standards, this analysis estimated conditional associations of single traffic-related pollutants. In addition, the exposure was assigned based on measurements from an air pollution monitoring station within 2 miles of a residence zip code.

The results of previous studies have been more consistent for low birth weight than for preterm birth, yet uncertainty remains as to the most vulnerable time in pregnancy with respect to exposure to ambient air pollution. Similarly, associations have been found with all of the criteria pollutants but little consistency as to which ones are the most critical. This variability in results may be due to the differences in the pollution sources and profiles in the locations where these studies have been conducted as well as differences in susceptibility across populations. For methodological reasons, described in Chapter 1, pollutants typically are evaluated individually; however, exposure to pollutant mixtures vary in different locations based on sources, topography and meteorology (HEI 2010).

Metrics based on exposure to traffic have been used as a surrogate for exposure to the mixture of pollutants from mobile sources. Proximity to traffic has been used most frequently to assess exposure (Wilhelm and Ritz 2003; Slama, Morgenstern et al. 2007; Genereux, Auger et al. 2008; Zeka, Melly et al. 2008). As reported in Chapter 5, results have demonstrated associations between high exposure to CO and NO_2 during pregnancy and deficits in spirometric measures later in childhood. These associations among these two correlated traffic-generated pollutants ($r=0.78$) suggest that chemicals in motor vehicle exhaust may affect fetal development.

In the same study population in Los Angeles, mentioned above, Wilhelm reported an increase in preterm birth for those exposed to higher levels of traffic, with higher associations among those with low socioeconomic status in winter months (Wilhelm and Ritz 2003). Traffic was defined as categorical values of distance weighted traffic density (DWTD) consisting of annual average daily traffic counts ($<20^{th}$ percentile, 20-80^{th} percentile, and $>80^{th}$ percentile). Socioeconomic status was assessed by unemployment, family poverty and income from public assistance obtained from the 1990 Census at the county level (Wilhelm and Ritz 2003). This study assessed exposure to a mixture of pollutants from traffic, a key source of ambient air pollution. Compared to simple traffic metrics, such as the distance to the nearest big road, the traffic density metric has the advantage of accounting for multiple roads surrounding a location of interest. The results are conditional on socioeconomic status and season of birth, rather than a marginal estimate, making it difficult to apply to the population level. Chapter 4 has a more detailed discussion of the value of marginal estimates.

Term low birth weight and preterm birth are two distinct adverse birth outcomes. The contrast between term low birth weight and preterm low birth weight is an issue of growth rather than a shortened duration of pregnancy. Simply put, term low birth weight occurs when the fetus grows more slowly during gestation, but the length of gestation is unaffected. In contrast, preterm birth occurs when the gestational period is shortened which can often result in low birth weight. The fetus may or may not have achieved its expected weight during those weeks of gestation. These distinct outcomes of low birth weight may have diverse etiologies and health consequences, but share many common risk factors: maternal age (<18, >35), height and weight, race and ethnicity (particularly African-Americans), single marital status, low socioeconomic status, cigarette smoking, alcohol consumption, drug use, parity, previous LBW or PTB, low weight gain during pregnancy, lack of prenatal care, hypertension, diabetes, infection, malnutrition; placental factors; or fetal factors including chromosomal abnormalities, genetic defects, growth hormone deficiency or short stature syndromes (Lee, Ha et al. 2003). Some of these factors, in particular low socioeconomic status and race have been shown to be associated with high exposure to air pollution levels (Zeka, Melly et al. 2008) and, therefore, may be important confounders in the relationship between air pollution exposure and adverse birth outcomes. The degrees to which these confounders affect estimation depend on the study population and quality of the data.

The majority of studies on air pollution and birth outcomes have used data from birth certificates due to the ready availability of these data and the need for large sample sizes. It has been demonstrated that in some geographic areas, factors that may confound the relationship between traffic-related air pollution and birth outcomes are not available from birth certificate

data (Salam, Millstein et al. 2005). However, data from a study in L.A. suggest the birth certificate adequately captured the most critical confounders for that population. Ritz, *et al.* obtained survey data from a sub-sample of a cohort of 58,316 births in 2003 in Los Angeles County, California. Based on these data, they found that trimester-specific effect estimates for air pollution exposure did not appear to be confounded by covariates not routinely collected on birth certificates such as occupation, income, maternal smoking and exposure to secondhand smoke and alcohol consumption (Ritz, Wilhelm et al. 2007; Ritz and Wilhelm 2008). While adjustment for covariates reported on birth certificates had the strongest influence on the pollutant effect estimates (up to 17%), additional adjustment for a large number of survey covariates (household income, regular source of prenatal care, consumed alcohol during pregnancy, smoked during pregnancy, lived in a house with a smoker during pregnancy, marital status, occupation, maternal weight gain during pregnancy) changed the effect estimates by less than 5%. These results support the feasibility of using birth certificates to control for confounding at the individual level.

A few prospective cohort studies have been carried out to address this study question. Prospective cohorts do not rely on birth certificates for covariate data and have the ability to collect additional data on important covariates such as residential history during pregnancy, maternal smoking and socioeconomic status. One large cohort study in Vancouver Canada, found a 26% increase in small for gestational age (SGA) and an 11% increase in LBW associated with residence within 50 meters of a highway. Individual air pollutants (NO_2, PM_{10}, $PM_{2.5}$, CO, SO_2) were also associated with SGA as well despite the low levels of ambient air pollution (Brauer, Lencar et al. 2008). A smaller prospective cohort study in the Netherlands found no association between residential proximity to traffic (measured as traffic density within a 150m buffer and proximity to a major road) and birth outcomes (SGA, LBW, PTB) (van den Hooven, Jaddoe et al. 2009). There are some disadvantages to prospective studies as well. They are more expensive to conduct and, therefore, are limited to the number of participants. Furthermore, if women are enrolled after pregnancy, recall bias may influence the results.

Several studies have included neighborhood-level socioeconomic status (SES) as a potential effect modifier and/or confounder (Wilhelm and Ritz 2003; Ponce, Hoggatt et al. 2005; Genereux, Auger et al. 2008; Zeka, Melly et al. 2008) in the relationship between traffic exposures and low birth weight. Wilhelm *et al.* found an association between traffic-related air pollution and preterm birth in low and middle-income neighborhoods, but not in the higher-income neighborhoods (Wilhelm and Ritz 2003). Ponce, *et al.* found that low SES neighborhoods were disproportionately affected by traffic-related air pollution exposure in the winter (Ponce, Hoggatt et al. 2005). Low SES neighborhoods were defined by census tracts meeting all three of the following criteria: greater than 10% unemployment, greater than 20% of families in poverty, and greater than 15% of individuals receiving public assistance (Ponce, Hoggatt et al. 2005).

The epidemiology of air pollution and birth outcomes is still evolving. If a relationship exists, it is subtle and not easy to characterize due to many methodological challenges (Ritz and Wilhelm 2008; Stillerman, Mattison et al. 2008). In summary, studies have used a variety of exposure assessment methods to characterize ambient air pollution in diverse geographic

locations. The majority of the data on outcomes and covariates have been from birth certificates alone; however a few prospective cohort studies have been conducted and have shown mixed results. Previous studies have used traditional regression methods to estimate conditional association. This chapter explores the use of semi-parametric causal inference methods to estimate associations between exposure to traffic density during pregnancy and term low birth weight. I restrict this chapter to one exposure metric and one outcome to demonstrate the various causal inference methods rather than fully explore the range of subject-matter hypotheses I plan to evaluate in future work. Chapter 4 has a more thorough discussion of the advantages of these statistical methods. The traffic density metric was chosen because it incorporates traffic counts rather than a simple distance to roadway or length of roadways near one's residence. Traffic density refers to the number of cars that travel over each length of road on average during an interval of time. The rationale for starting with term low birth weight is that the data on birth weight, as opposed to gestational age, are the most reliable. Although results of previous studies have been inconsistent, biological plausibility, discussed in Chapter 3.5, suggest that air pollution exposure could potentially affect fetal growth.

6.2 Methods

6.2.1 Study Population

The Study of Air pollution, Genetics and Early life events (SAGE) was funded by the National Institute for Environmental Health Science (NIEHS) to investigate causal associations between exposure to traffic-related ambient air pollution during pregnancy and birth outcomes. Two sources of exposure data were linked to California birth certificate data from four counties in the San Joaquin Valley (Kern, Fresno, San Joaquin and Stanislaus) to evaluate the impact of traffic-related air pollution exposures during pregnancy on adverse birth outcomes in SAGE. Exposure data included detailed traffic proximity data based on geo-coded maternal residences as well as highly refined exposure data from California Regional PM_{10}/ $PM_{2.5}$ Air Quality Study (CRPAQS). This research was approved by the University of California, Berkeley Office for Protection of Human Subjects and the California State Committee for the Protection of Human Subjects.

6.2.2 Outcome Ascertainment

Birth certificates, provided by the California Department of Health Services (Sacramento, CA) were used to identify subjects. The birth certificate data collection and quality control is the responsibility of Office of Vital Records, which maintains a permanent public record of each birth, death, and fetal death that has occurred in California since 1905. According to the Office of Health Information and Research, no more than 1-2% of the birth data collected in California lack information on such important factors as gestational age, birth weight, maternal age, infant sex, maternal race, prenatal care information, and maternal education (OHIR 2010).

Gestational age is most often calculated based on the mothers' reported last menstrual period (LMP); however, this calculation can be biased due to imperfect recall of the LMP or if the mother's menstrual cycle is irregular and the luteal phase is longer than average. I chose to use "term low birth weight" as the outcome of interest because there are greater concerns

about the data quality of gestational age. Low birth weight is defined at <2500 grams and term was defined as ≥ 37 weeks gestation. There were a number of exclusions used to isolate term low birth weight as the outcome and not to include other adverse birth outcomes. Infants with gestation >44 weeks were excluded due to likely data quality concerns for LMP. Infants with birth weight <1000 grams or >5000 grams also were excluded because of likelihood of complications such as birth defects, maternal diabetes or preterm birth as well as data quality issues related to the validity of the weight measurement. Finally, mothers with pregnancy complications such as hypertension, diabetes or uterine bleeding were excluded based on the assumption that the potential effects of traffic exposure would be far outweighed by the influence of these maternal conditions (Ritz 1999). It should be noted that information on pregnancy complications is known to be incomplete and therefore my exclusion of the 1.5% of mothers with these complications is not likely to not identify all possible cases of adverse birth outcomes due to these conditions.

6.2.3 Covariates

The list of covariates considered in this study started with variables from the birth certificate that were used in previous studies mentioned above. The validity and reliability of certain characteristics vary. The variables that I included in this analysis from the birth certificate include: age of mother (<20, 20-35, >35), race of mother (White, Hispanic, African-American, Asian, other), education level of mother (no high school, some high school, some college, bachelors or other degree), parity (0, ≥ 1), prenatal care (initiated in first, second or third trimester), paid birth expenses with Medi-Cal, sex of infant, year (2000-2006) and county of mother's residence (Fresno, Kern, Stanislaus, San Joaquin). Although the birth certificate does provide additional information, I did limit the candidates to variables associated with the exposure and/or outcome. Selection of these variables into the models as covariates will be discussed in the results below.

As mentioned above, socioeconomic measures, such as prevalence of poverty and unemployment, have been associated with adverse birth outcomes (Ponce 2005). To obtain a more complete adjustment for confounding, I did not rely on the sole indicatory of SES on the birth certificate (birth costs paid by Medi-Cal). Instead, as Ponce did in a previous study, I created an indicator variable for low SES, which was defined as block group level unemployment $>10\%$, income from public assistance $>15\%$ and families below poverty level $>20\%$ in the 2000 U.S. Census at the block group level (Census 2000). This variable may not pertain directly to a SAGE individual, but is meant to provide contextual information about the neighborhoods in which the SAGE study population lived.

6.2.4 Exposure Assessment

All 2000-2006 births to women living in four counties in the San Joaquin Valley of California (Fresno, Kern, Stanislaus and San Joaquin) were identified. The home locations were geo-coded with Arc GIS software (Environmental Systems Research Institute, Redlands, CA). Addresses were corrected with ZP4 software (Semaphore Corporation, Aptos, CA) in ArcView and SAS (SAS Institute Inc., Cary, NC). The geo-codes were sent to Sonoma Technology Inc., Sonoma, CA (STI) for estimates of traffic and air pollution exposure during the time of pregnancy. STI provided three metrics to characterize traffic for the SAGE study population within the San Joaquin Valley Air Basin (SJVAB). Two metrics represent the

proximity of a subject's residence to roads of various sizes and lengths. A third traffic metric, traffic density described in more detail below, represents both proximity to and traffic volume on, local roads.

As noted above, single pollutant models can be problematic due to variations in pollution profiles and sources across communities and seasons. This analysis, therefore, is focused on traffic metrics instead. In the published literature, traffic has been measured in a number of different ways. One commonly used metric is distance to nearest roads. Experimental studies and dispersion theory indicate that pollution levels generally decrease inversely with distance downwind from roadways. Thus, distance (or inverse distance) to various types of roadways is a potential indicator of exposure to traffic-related pollution. These distances are often specified by road classes 1-4 and assigned respectively to freeways, primary highways, secondary highways, and local roads, but generally ignore the effect of wind direction on the likely differences in decline in pollutant concentrations for upwind and downwind residences. An additional traffic metric is length of roads in a 200m radius buffer around each location of interest. This metric also is specified by road class (1-4). This proximity metric may represent the density of roads in an area better than the distance to the nearest road; however, effects of wind direction still are ignored. The metric used in this analysis is a dimensionless indicator of traffic density based on distance-decayed annual average daily traffic (AADT) volumes. Traffic density is calculated using roadway link-based traffic volumes which are derived from traffic count data and has been used in other health effects studies (Kan, Heiss et al. 2008). Density plots are generated within a GIS using a linear decay function that approximates the fall-off of ambient concentrations with increasing distance away from roadways. An advantage to the traffic density parameter is that it accounts for the combined influence of all roadways and activity (for which data exist) near each location, but not for wind direction effects (Penfold 2009).

The roadway location data were obtained from Tele Atlas/Geographic Data Technology (GDT) Dynamap in 2005. Traffic density maps were created using one parameterization for dispersion, in which density decreases by 90% at 300 m from the value at the edge of the roadway, which is consistent with data from numerous dispersion studies (Zhu, Hinds et al. 2002; Zhu, Kuhn et al. 2006). GIS tools were used to extract the traffic densities from the map at the location s of the residences of the study population. Overall, with this method, volumes were assigned to 93% of class 1 roads, 88% of class 2 roads, 65% of class 3 roads, and 7% of class 4 roads. Since, the GDT traffic count data were mostly for 1995-2000 and the period of interest was 2000-2006, the counts were scaled up to represent 2003 traffic based on traffic based on county average vehicles-miles-traveled (VMT) growth (CADoT 2004).

The assignment of traffic count to links is straightforward for interstate freeways and other high-volume roads were count data are available for almost every link. On moderate and smaller roads, traffic count data are generally sparse, and imputation of link volumes is required. An extrapolation method based on roadway name, connectivity and distance was used to assign traffic count data to roadway links. Links were connected up to 5km, 7km, and 10km from the traffic count locations for road Classes 4, 3, 2 and 1, respectively. Links with like names and within the specified distance were only assigned traffic count data when the

links were connected. This extrapolation method produces consistent assignments of traffic volumes that have few gaps on the named roadways with count data; however, smaller local roads lacking count data are not included (Penfold 2009).

6.2.5 Statistical Methods

G-computation and targeted maximum likelihood estimation (TMLE) are used to estimate the counterfactual predicted probabilities of low birth weight had the entire study population been exposed to each quartile of traffic density. As described in more detail in Chapter 4, these causal inference methods are based on counterfactual theory. Briefly, counterfactuals are the set of possible outcomes that would be observed under each possible treatment, if, contrary to fact, each person could be observed after exposure to each level of the treatment (traffic density exposure). The goal of this analysis is to estimate, at the population level, the predicted probability of term low birth weight had everyone been exposed to each quartile of traffic density, which can be written as

$$E_w\{E(Y \mid A=a, W)\}.$$

where Y is the outcome of term LBW, A is quartile of traffic density exposure during pregnancy, and W is the vector of covariates.

The TML estimation consists of two modeling steps. The first step is to use G-computation, an imputation technique, to create the full data (*i.e.,* four predicted outcomes as if each child had been exposed to each quartile of traffic density during the prenatal period).

$$Q^0(W, A) = P(Y \mid A, W).$$

The second step begins with modeling the treatment (exposure) mechanism and in the case where A is binary, can be defined as:

$$g(A, W) = P(A \mid W).$$

The most straightforward way to employ a TML estimator is with an indicator of treatment (rather than use of a continuous variable). Therefore, I created four indicators for each quartile of traffic density. Both of these nuisance models were selected by the Deletion/Substitution/Addition (D/S/A) algorithm. Once $g(A,W)$ was specified, I fitted a logistic regression model for the probability of each quartile of traffic density exposure given variables selected from the D/S/A. I calculated the clever covariate, h, with coefficients from this treatment model:

$$h(A, W) = I(A=1)/(g(1 \mid W) - I(A=0)/g(0 \mid W).$$

The clever covariate, h, serves as a weight to "target" the parameter of interest in the final model. The clever covariate is added to the Q^0 model to adjust the parameter of interest by incorporating the treatment mechanism and to produce and updated Q^1 model and $\hat{\varepsilon}$ is estimated using logistic regression with the logit[Q^0(A,W)] as an offset:

$$\hat{Q}^1(W, A) = \hat{Q}^0 + \hat{\varepsilon} * h(A, W).$$

In summary, the primary parameter of interest is:

$$\hat{\theta}^{TMLE} = \hat{E}_W[\hat{Q}^1(W, a)]$$

I specified that the models should be restricted to a size of 10, second order interactions, and a maximum sum of powers of 3—*i.e.*, up to 10 variables in the model, two-way interaction terms between any of the variables and all the variables up to second order polynomials.

I calculated the standard errors to obtain proper confidence intervals for inference on the TMLE estimates with an influence curve:

$$IC(Y,W \mid \hat{Q},\hat{g},\hat{\theta}) = \frac{I(A=a)}{\hat{g}(1 \mid W)}(Y - \hat{Q}(a,W)) + \hat{Q}(a,W) - \hat{\theta}$$

The influence curve allows one to calculate how much each observation deviates from the estimate in order to infer the variance of that estimator. I calculated confidence intervals at the 95% level.

I included a population intervention model (PIM) to show an extension of the TML estimate. However, since it is not the primary parameter of interest, the statistical inference is not reported here and further exploration of this parameter will continue in future work. As noted in Chapter 4, a PIM compares the mean outcome in a population under one treatment-specific counterfactual to the mean outcome in the observed population:
$E[Y] - E[Y_a] = E(Y) - E_W(Y \mid A=a, W)$ under assumptions.

For comparison purposes, I performed traditional conditional analyses similar to previous studies. I used logistic regression to estimate the odds of LBW among those in each quartile of exposure, using the lowest quartile as a reference. These analyses included both crude associations as well as adjusted associations conditional on all the candidate covariates. Analyses were performed using R software (R Foundation for Statistical Computing, Vienna Austria, version 2.10.1, package: DSA, version 3.1.1).

6.3 Results

Of the 316,110 births in San Joaquin Valley during 2000-2006, the SAGE study population was restricted to singleton births (n=8,387 multiples). Additionally, infants with birth weight less than 1000 grams or greater than 5000 grams and gestational length missing, less than 20 weeks or more than 44 weeks also were excluded (n=34,539). The study population was restricted further to gestation between 37 and 44 weeks to capture *term* low birth weight rather than preterm low birth weight (n=31,391 excluded). Births with the maternal conditions during pregnancy such as hypertension, diabetes or uterine bleeding were excluded (n=4,762). The final study population (n=237,031) is described in Table 6.1.

Table 6.2 shows the distribution of covariates for low birth weight and normal weight infants among full term pregnancies. The proportion of low birth weight infants was higher in women who were less than 20 years of age, women who were of African-American race and for those whose birth costs were paid by Medi-Cal. The distribution of covariates across quartiles of exposure to traffic density during pregnancy is in Table 6.3. High traffic exposure is associated with maternal African-American race, maternal education (no college) and low socioeconomic status (measured at the neighborhood level according to the 2000 U.S. census and at the individual level by proxy of birth costs paid by Medi-Cal). The county of Fresno comprises the largest portion of the second and fourth quartile of exposure. As shown in Table 6.4, Fresno and San Joaquin had the greatest mean traffic density compared to the other Kern and Stanislaus counties.

6.3.1 Targeted Maximum Likelihood Estimation

The variables that were selected by the D/S/A in both the outcome (Q) and exposure (g) models are listed in Table 6.5. Four of the variables were predictive of both the exposure

and the outcome: maternal age >35, African-American race, education and the trimester of initiation of prenatal care. Additional variables came into each of the nuisance models. Table 6.6 shows the predicted probabilities of term low birth weight had everyone been exposed to each quartile of traffic density. If everyone were exposed to the highest quartile of traffic density during pregnancy, there would be a 2.27% probability of term low birth weight in the population, compared to the 2.02% of term low birth weight had everyone been exposed to the lowest quartile of traffic density during pregnancy. These results were estimated by TMLE and the statistical inference was calculated using the influence curve. The results show that the estimated probabilities of low birth weight are lower in the first and third quartile and higher in the second and fourth quartile. The highest quartile of traffic density exposure is associated with significantly higher term low birth weight compared to the lowest quartile; however, there is not a monotonic exposure-response relation. Figure 6.1 is a graphical representation of the TMLE results.

6.3.2 Population Intervention Model

Based on the PIM estimates (Table 6.7), if a population invention could reduce everyone's traffic density exposure during pregnancy to that of the lowest quartile, it is estimated that more infants would be born of normal weight rather than low birth weight. In other words, there would be a reduction of 6% (0.0014/0.0216=0.0648) of term low birth weight. Although statistical inference is not presented for this model, this PIM is simply an example an extension of TMLE to alternative parameters of interest.

6.3.3 Traditional Regression Estimation

The results based on more traditional methods of analysis (Table 6.8) provide the *conditional* estimates of term low birth weight *among* those in each quartile of exposure. This method imposes an arbitrary model to the estimation of its parameter. According to the logistic regression, the odds of term low birth weight is 10% higher among births where the mother was exposed during pregnancy to the highest quartile of traffic density compared to the lowest quartile of traffic density, conditional on maternal age, race, education, county of maternal residence, birth costs paid by Medi-Cal, low SES, parity, sex of infant, initiation of prenatal care, year of birth.

6.4 Discussion

The results from the TMLE analysis show a difference in the probability of term low birth weight had everyone been exposed to different quartiles of traffic density exposure during pregnancy. The results do not show a clear exposure-response across the quartiles; however, there is a significant difference in the predicted probability of low birth weight between the highest and lowest quartile of exposure, showing that higher traffic density is associated with increased probability of low birth weight.

The TMLE methods assume that there are no violations of the experimental treatment assignment (ETA) or positivity assumption, no unmeasured confounding, consistency, temporal ordering and correct model specification as described in more detail in Chapter 4.2.4. The ETA assumption was tested by plotting the probability of being treated versus the log odds to show the distribution of probabilities of treatment (Figure 6.3). The probability of treatment is not 0 or 1 for any observation, therefore those who were exposed to specific

quartiles of traffic density were not deterministic based on the treatment model. The series of points at the 0 and 1 level on the y-axis show the distribution of observations that were in the each quartile (at 1) and not in each quartile (at 0) and their corresponding log odds. These plots indicate that there is no ETA violation in these data. Although the plot of the second quartile appears to have fewer observations, this is not the case. Rather, there were fewer combinations of covariate patterns in the treatment model which predicted whether or not someone was exposed to that level of traffic density. The no unobserved confounding (NUC) assumption requires that all confounding factors are measured and included in the modeling steps. The consistency assumption requires that the observed outcome is a member of the set of all possible outcomes. The NUC and consistency assumptions are not testable. Temporal ordering must be maintained; that is, the confounders must precede treatment, which must precede the outcome. In this study, the confounders and treatment, which at times coincide, both precede the outcome. The assumption for model specification cannot be tested; however the model fitting was optimized with the use of data-adaptive algorithms for modeling the nuisance parameters.

In contrast to previous studies mentioned above, this study uses detailed traffic density exposure data during the entire pregnancy. This exposure assessment does not target specific periods of pregnancy, rather, it assumes the exposure is constant across the entire pregnancy. Therefore, the traffic metric does not account for seasonal differences throughout the year, which may correspond to certain periods of gestation. The density of traffic across different seasons may vary by area and the chemical and physical transformations of the vehicle exhaust certainly vary by season. More on this topic is discussed in Chapter 1. In contrast to single air pollutant models, the exposure assessment estimates exposure to a source instead of the measured ambient pollutants. The location of residence is also more precise by geo-coding the street address as opposed to some previous studies which used only the zip code of the mothers' residence and assigned them to a monitoring station within 2 miles (Ritz 1999). Although traffic density captures a mixture of several air pollutants, it is not able to capture complete profiles of any given pollutant as can air pollution monitoring systems. The use of the traffic density metric does present a trade-off between examining season variation and effects during specific trimester versus a more inclusive exposure metric of a key source emitting a mixture of pollutants. Future analyses of the SAGE population will incorporate highly refined exposure data from California Regional PM_{10}/ $PM_{2.5}$ Air Quality Study (CRPAQS) to examine the influence of traffic in conjunction with air pollutants on birth outcomes. Further investigation into the individual pollutants may explain the inconsistent pattern in the second and third quartiles of traffic density in relation to term low birth weight or a chance finding.

This analysis was limited to the outcome of term low birth weight. The study population was restricted to full term births to identify a specific etiologic occurrence. The birth certificate includes additional data on preterm birth and small for gestational age. Future analyses will also address the associations of both traffic density and ambient air pollutant levels in relation to other birth outcomes using causal inference methods.

The SAGE study population is large with a broad geographic area, characterized by areas of high ambient air pollution. Most studies of traffic exposure during pregnancy and

birth outcomes have included only single metropolitan areas. Although this diversity increases the variability and potential confounding, with such a large sample size, this wide geographic area provides an opportunity to estimate this relationship across a large population with a wider gradient of exposure.

In addition to birth certificate characteristics, this study included neighborhood socioeconomic status from the U.S. census data. Neighborhood SES was selected by the D/S/A in the prediction model of exposure, suggesting that SES influenced the level of traffic a mother was exposed to. This suggests that perhaps additional variables from the U.S. census or other public records could help characterize the contextual components of neighborhoods in future analyses. Neighborhood level socio-demographic factors such as violence, deprivation, nutrition, and alcohol-use will be used in future analyses to explore potential confounding and/or effect modification.

Other studies (Ritz, Yu et al. 2000; Ritz, Yu et al. 2002) have incorporated census-level metrics as surrogates for individual-level exposures to these risk factors. Although they are likely to be an improvement over crude analyses, census-level metrics have been criticized for several reasons. The census geographic unit has several limitations, specifically: population changes, the assumption of uniform distribution, and fixed boundaries that define all residents as having the same aggregate characteristics within the unit. For some residents, a census geographic unit does not capture adequately important details of their neighborhood (Coulton, Korbin et al. 2001).

There were certain variables that were selected to both the outcome and treatment model, suggesting they influence both the level of exposure and the birth outcome. The trimester at which prenatal care was initiated, maternal age, race, education and SES at the individual level (measured by whether or not birth costs were paid by Medi-Cal) were associated with both exposure and outcome and therefore, were likely confounders in the relationship between prenatal exposure to traffic and term low birth weight. In this study, the change in estimates between the crude and adjusted analyses (and observed and predicted), both for the traditional and causal inference methods, suggest there is confounding by these variables, and therefore, methods for proper adjustment of confounding are required.

Further exploration of effect modification by county in this dataset will be explored in future analyses. The contrast between Stanislaus and Fresno counties, in particular, may help explain the unclear results for the second and third quartiles of exposure. There may be qualitative differences in the traffic mixture and other pollutants in the two counties. Future studies, mentioned above, that incorporates ambient air pollutant exposure data will be able to investigate this further.

Finally, TMLE allows the estimation of a simple measure of association that assumes no particular model for the regression. In this example, this semi-parametric risk estimate is not just a byproduct from an arbitrary regression model. These methods allow use of machine learning algorithms to optimize fit and reduce bias. Empirical standard errors are available because the estimate is equivalent to a simple estimating equation approach. In addition to providing a marginal estimate, TMLE accounts for heterogeneity of the individual by targeting the parameter of interest. In this study, the traditional methods provided a similar overall pattern in the relationship between exposure to traffic density during pregnancy and

term low birth weight, though the results are not directly comparable and have different interpretations.

In summary, the results from all of these analyses demonstrate there is an association between traffic density exposure during pregnancy and term low birth weight. Although the relationship is not consistent across each quartile, the probability of term LBW is significantly lower in the first quartile relative to the fourth quartile. As with all observational studies, these results assume there is no unmeasured confounding. In additional follow-up work, I intend to explore the influence of the physical and social context in which these associations have been observed.

6.5 List of Tables and Figures

Table 6.1. Characteristics of SAGE population.

Covariates	**N** (N=237,031)	(%)
Maternal age (years)		
<20	32,270	(13.6)
20-35	179,816	(75.9)
>35	24,942	(10.5)
Maternal race		
Asian	17,738	(7.5)
African-American	11,560	(4.9)
Hispanic	132,605	(55.9)
White	71,522	(30.2)
Other	3,606	(1.5)
Maternal education		
No high school	28,027	(11.8)
Some high school	124,128	(52.4)
Some college	49,412	(20.8)
Bachelor's or other degree	30,090	(12.7)
Missing	5,374	(2.3)
Birth costs paid by Medi-Cal		
Yes	127,564	(53.8)
No	109,467	(46.2)
Low socioeconomic status*		
Yes	41,745	(17.6)
No	195,286	(82.4)
Parity		
0	83,819	(35.4)
>=1	153,212	(64.6)
Sex of infant		
Male	120,456	(50.8)
Female	116,575	(49.2)
Initiation of prenatal care		
First trimester	192,905	(81.4)
Second trimester	32,676	(13.8)
Third trimester	7,317	(3.2)
Unknown	4,133	(1.7)
Year of birth		
2000	30,788	(13.0)
2001	31,707	(13.4)
2002	32,534	(13.7)

2003	33,082	(14.0)
2004	34,331	(14.5)
2005	35,567	(15.0)
2006	39,022	(16.5)
County of maternal residence		
Fresno	77,093	(32.6)
Kern	56,318	(23.8)
San Joaquin	59,680	(25.2)
Stanislaus	43,940	(18.5)

* Low socioeconomic status was defined as block group level unemployment >10%, income from public assistance >15% and families below poverty level >20% at the block group level from the 2000 census.

Table 6.2. Number (column percent) of study population in each outcome group by demographic characteristics.

Covariates	Term low birth weight N=5,123 (2.2%)		Term normal weight N=231,908 (97.8%)	
	N	(%)	N	(%)
Maternal age (years)				
<20	968	(18.9)	31,302	(13.5)
20-35	3,561	(69.5)	176,258	(76.0)
>35	594	(11.6)	24,348	(10.5)
Maternal race/ethnicity				
Asian	584	(11.4)	17,154	(7.4)
African-American	527	(10.3)	11,033	(4.8)
Hispanic	2,653	(518)	129,952	(56.0)
White	1,261	(24.6)	70,261	(30.3)
Other	98	(1.9)	3,508	(1.5)
Maternal education				
No high school	539	(10.5)	27,488	(11.9)
Some high school	3,021	(59.0)	121,107	(52.2)
Some college	975	(19.0)	48,437	(20.9)
Bachelor's or other degree	452	(8.8)	29,638	(12.8)
Missing	136	(2.7)	5,238	(2.3)
Birth costs paid by Medi-Cal				
Yes	3,110	(60.7)	124,454	(53.7)
No	2,013	(39.3)	107,454	(46.3)
Low socioeconomic status*				
Yes	1,102	(21.5)	40,643	(17.5)
No	4,021	(78.5)	191,265	(82.5)
Parity				
0	2,303	(45.0)	81,516	(35.1)
>=1	2,820	(55.0)	150,392	(64.9)
Sex of infant				
Male	2,221	(43.4)	118,235	(51.0)
Female	2,902	(56.6)	113,673	(49.0)
Initiation of prenatal care				
First trimester	3,887	(75.9)	189,018	(81.5)
Second trimester	870	(17.0)	31,806	(13.7)
Third trimester	193	(3.8)	7124	(3.1)
Unknown	173	(3.4)	3960	(1.7)
Year of birth				

2000	608	(11.9)	30,180	(13.0)
2001	675	(13.2)	31,032	(13.4)
2002	688	(13.4)	31,365	(13.7)
2003	717	(14.0)	32,813	(14.0)
2004	700	(13.7)	33,631	(14.5)
2005	816	(15.9)	34,751	(15.0)
2006	919	(17.9)	38,103	(16.3)
County of maternal residence				
Fresno	1,757	(34.3)	75,336	(32.5)
Kern	1,253	(24.5)	55,065	(23.7)
San Joaquin	1,293	(25.2)	58,387	(25.2)
Stanislaus	820	(16.0)	43,120	(18.6)

* Low socioeconomic status was defined as block group level unemployment >10%, income from public assistance >15% and families below poverty level >20% at the block group level from the 2000 census.

Table 6.3. Number (column percent) of study population in each exposure quartile by demographic characteristics.

Covariates	1st Quartile Traffic Density (N=59,197)		2nd Quartile Traffic Density (N=59,271)		3rd Quartile Traffic Density (N=59,210)		4th Quartile Traffic Density (N=59,353)	
	N	(%)	N	(%)	N	(%)	N	(%)
Low Birth Weight	1149	(1.94)	1322	(2.23)	1232	(2.09)	1420	(2.39)
Maternal age								
<20	6,878	(11.6)	7,845	(13.2)	8,287	(14.0)	9,260	(15.6)
20-35	45,014	(76.1)	44,706	(75.5)	45,015	(76.0)	45084	(76.0)
>35	7,305	(12.3)	6,720	(11.3)	5,908	(10.0)	5,009	(8.4)
Maternal race/ethnicity								
Asian	4,582	(7.8)	4,135	(7.0)	4,006	(6.8)	5,012	(8.4)
African-American	1,826	(3.1)	2,511	(4.2)	3,003	(5.1)	4,220	(7.1)
Hispanic	32,054	(54.2)	32,880	(55.5)	33,892	(57.2)	33,779	(56.9)
White	19,631	(33.2)	18,837	(31.8)	17,553	(29.7)	15,501	(26.1)
Other	1,101	(1.9)	908	(1.5)	756	(1.3)	841	(1.4)
Maternal education								
No high school	6,915	(11.7)	6,743	(11.4)	7,360	(12.4)	7,009	(11.8)
Some high school	28,352	(47.9)	29,826	(50.3)	31,722	(53.6)	34,228	(57.8)
Some college	12,898	(21.8)	12,670	(21.4)	11,989	(20.3)	11,855	(20.0)
Bachelor's or other degree	9,795	(16.6)	8,704	(14.7)	6,663	(11.3)	4,928	(8.3)
Missing	1,237	(2.1)	1,328	(2.2)	1,476	(2.5)	1,333	(2.3)
Birth costs paid by Medi-Cal								
Yes	27,547	(46.5)	30,053	(50.7)	33,024	(55.8)	36,940	(62.2)
No	31,650	(53.5)	29,218	(49.3)	26,186	(44.2)	22,413	(37.8)
Low socioeconomic status*								
Yes	5,654	(9.6)	9,065	(15.3)	11,084	(19.8)	15,942	(26.9)
No	53,543	(90.5)	50,206	(84.7)	48,126	(81.3)	43,411	(73.1)
First born – Parity								
0	20,468	(34.6)	20,910	(35.3)	20,809	(35.1)	21,632	(36.5)
>=1	38,729	(65.4)	38,361	(64.7)	38,401	(64.9)	37,721	(63.6)
Sex of infant								
Male	30,129	(50.9)	29,948	(50.5)	30,059	(50.8)	30,320	(51.1)
Female	29,068	(49.1)	29,323	(49.5)	29,151	(49.2)	29,033	(48.9)
Initiation of prenatal care								
First trimester	48,781	(82.4)	49,195	(83.0)	47,456	(80.2)	47,473	(80.0)

Second trimester	7,627	(12.9)	7,428	(12.5)	8,618	(14.6)	9,003	(15.2)
Third trimester	1,823	(3.1)	1,596	(2.7)	1,968	(3.3)	1,930	(3.3)
Unknown	966	(1.6)	1,052	(1.8)	1,168	(2.0)	947	(1.6)
Year of birth								
2000	7,194	(12.2)	7,766	(13.1)	7,879	(13.3)	7,949	(13.4)
2001	7,470	(12.6)	7,862	(13.3)	8,127	(13.7)	8,248	(13.9)
2002	7,762	(13.1)	8,208	(13.9)	8,276	(14.0)	8,288	(14.0)
2003	8,183	(13.8)	8,260	(13.9)	8,380	(14.2)	8,259	(14.0)
2004	8,762	(14.8)	8,597	(14.5)	8,489	(14.3)	8,483	(14.3)
2005	9,339	(15.8)	8,951	(15.1)	8,560	(14.5)	8,717	(14.7)
2006	10,487	(17.7)	9,627	(16.2)	9,499	(16.0)	9,409	(15.9)
County of maternal residence								
Fresno	13,910	(23.5)	21,155	(35.7)	16,303	(27.5)	24,725	(43.3)
Kern	15,698	(26.5)	16,488	(27.8)	14,574	(24.6)	9,558	(16.1)
San Joaquin	16,456	(27.8)	11,023	(18.6)	15,787	(26.7)	16,414	(27.7)
Stanislaus	13,133	(22.2)	10,605	(17.9)	12,546	(21.2)	7,656	(12.9)

* Low socioeconomic status was defined as block group level unemployment >10%, income from public assistance >15% and families below poverty level >20% at the block group level from the 2000 census.

Table 6.4. Traffic Density in San Joaquin Valley and by county.

County	N Births	Mean (SD)	Median	IQR	Maximum*
Total	237,031	35.1 (52.6)	16.5	45.1	554.5
Fresno	77,093	44.4 (59.6)	21.1	60.1	551.7
Kern	56,318	25.0 (37.6)	11.7	32.4	488.9
San Joaquin	59,680	39.6 (60.0)	20.1	50.5	532.6
Stanislaus	43,940	25.8 (39.9)	13.3	35.2	554.5

*Minimum=0, therefore, Maximum=Range

Table 6.5. Covariates chosen by the D/S/A for the Q (outcome=term LBW) and g (treatment=quartile of traffic density) model.

Covariates	Outcome Model (Q) G-computation	Treatment Model (g) TMLE
Maternal age (years)		
<20		
>35	X	X
Maternal race/ethnicity		
Asian	X	
African-American	X	X
Hispanic		X
White		
Other		
County of maternal residence		
Fresno		X
Kern		X
San Joaquin		X
Stanislaus		X
Maternal education	X	X
Birth costs paid by Medi-Cal	X	X
Low socioeconomic status*		X
Parity	X	
Sex of infant		
Initiation of prenatal care	X	X
Year of birth		X

* Low socioeconomic status was defined as block group level unemployment >10%, income from public assistance >15% and families below poverty level >20% at the block group level from the 2000 census.

Table 6.6. Observed percent of term low birth weight in SAGE population compared to the results from G-computation and TMLE analyses: percent predicted probability of term low birth weight had everyone been exposed to each quartile of traffic density.

	1st Quartile	**2nd Quartile**	**3rd Quartile**	**4th Quartile**
Observed	1.94	2.23	2.09	2.39
G-computation	2.02	2.28	2.07	2.26
TMLE (CI)	2.02 (1.90, 2.12)	2.29 (2.18, 2.40)	2.06 (1.96, 2.17)	2.27 (2.16, 2.38)

Table 6.7. Results from Population Intervention Model.

	1st Quartile	2nd Quartile	3rd Quartile	4th Quartile
PIM	0.0014	-0.0013	0.0010	-0.0011

Table 6.8. Results from traditional regression analyses: Odds of term low birth weight for each quartile of traffic density compared to the lowest quartile of exposure.

	2nd Quartile	**3rd Quartile**	**4th Quartile**
Crude	1.15	1.07	1.24
Adjusted*	1.11	1.02	1.10

*Adjusted for maternal age, race, education, county of maternal residence, birth costs paid by Medi-Cal, low SES, parity, sex of infant, initiation of prenatal care, year of birth (with no interactions).

Figure 6.1. TMLE estimates (and 95% confidence intervals) of the percent predicted probabilities of term low birth weight had everyone been exposed to each quartile of exposure to traffic density during pregnancy.

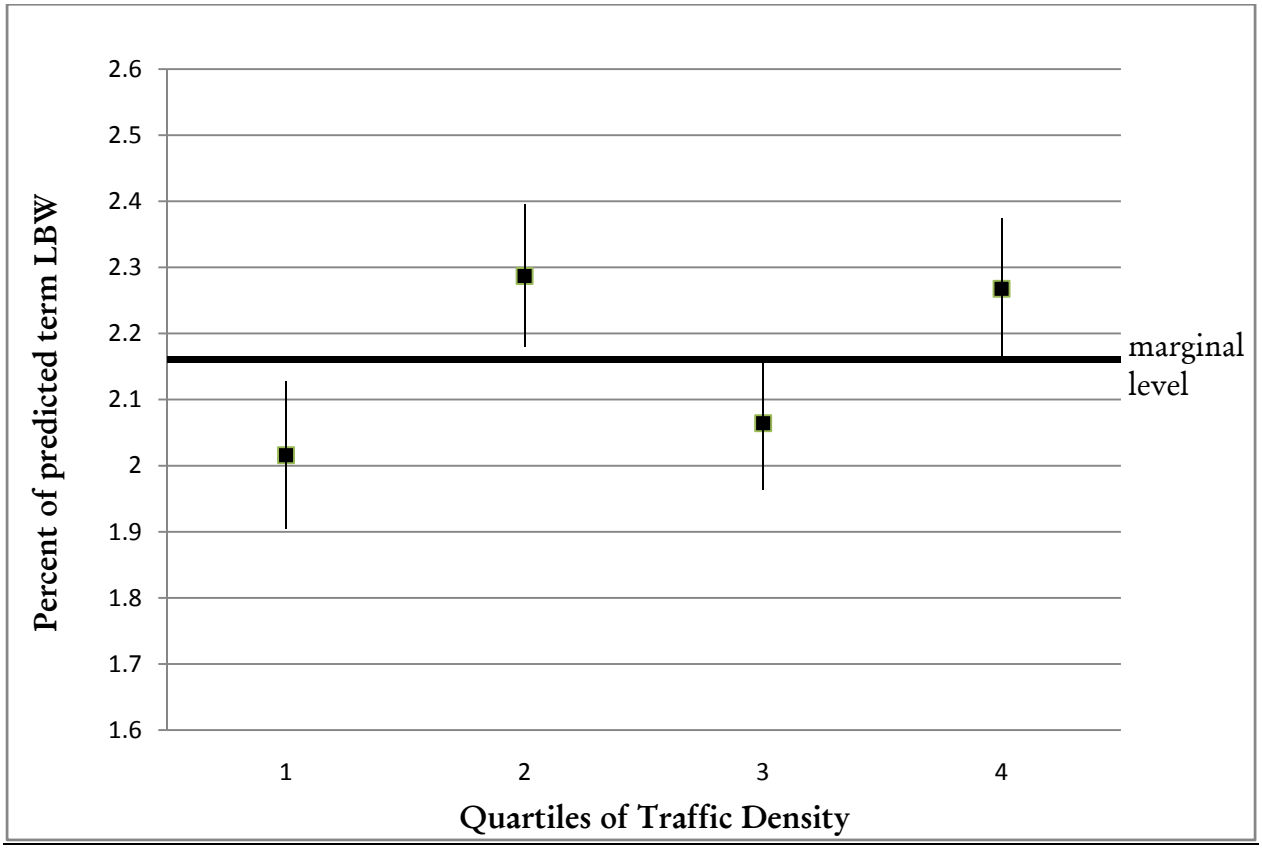


Figure 6.2. Annual average daily traffic volumes assignments to roadways. (Sonoma Technology Institute, Inc.)

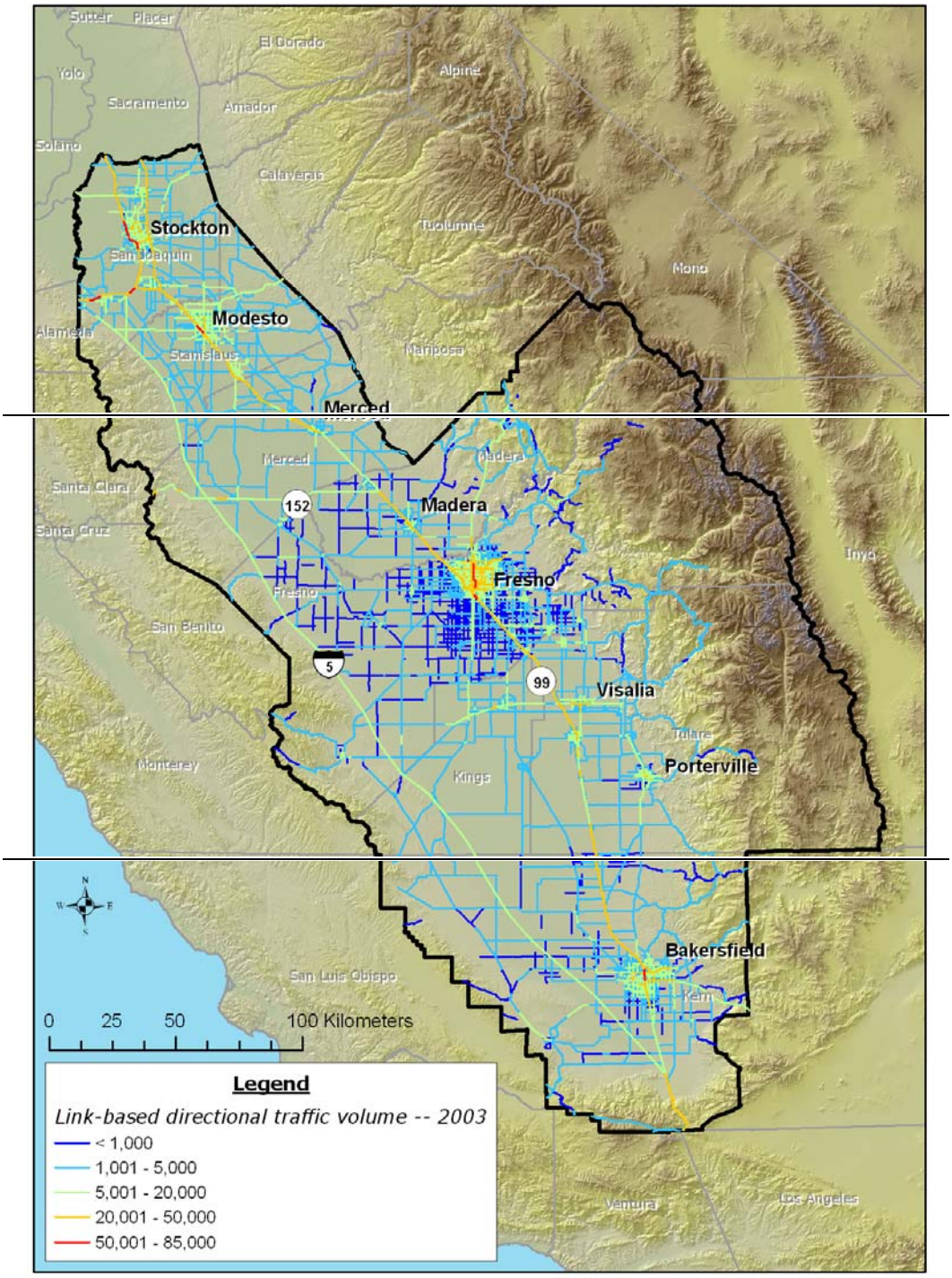
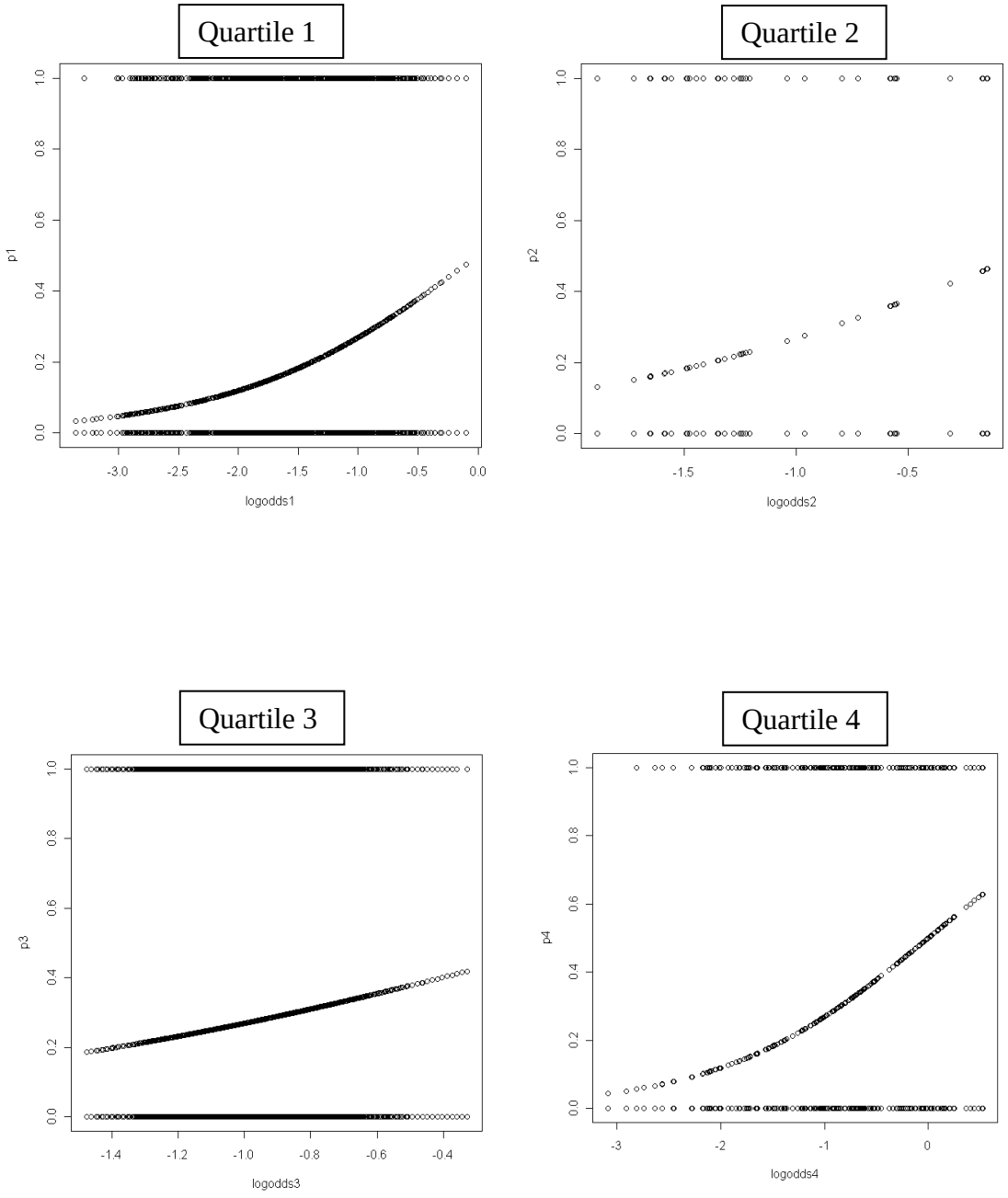


Figure 6.3. Plot of probability of treatment (exposure to traffic density) versus the log odds of treatment for each quartile of exposure.



Conclusion

The research presented in this dissertation addressed the relationship between traffic-related air pollution and two pressing child health outcomes, asthma and low birth weight. As described earlier, asthma is the leading cause of school absenteeism and morbidity. Birth weight is one of the most critical predictors of infant mortality. Traffic-related pollution is a ubiquitous and complex exposure that changes with the evolving built environment, population growth and technology across the world. Both studies targeted a critical period of exposure – during pregnancy – to assess its effects during fetal development. The two statistical analyses evaluated novel causal inference methods and contrasted the parameters of interest and interpretation of the results with traditional regression methods. In particular, targeted maximum likelihood estimation (TMLE) was used to estimate the counterfactual marginal effect of traffic-related air pollution exposure during pregnancy on pulmonary function and term low birth weight. In other words, the predicted outcomes were compared had everyone been exposed to specific levels of air pollution during pregnancy. These causal inference methods estimate easily interpretable parameters with important public health implications.

The analysis of the FACES-LITE study supports earlier work on the association of ambient air pollution exposure during pregnancy and lung function in children by using the repeated measures of lung function and applying causal inference methods. Furthermore, this is one of the first subject matter applications of the TMLE method. The study found that above median levels of ambient NO_2 exposure during the first and second trimesters were associated with deficits in pulmonary function for all of the age groups.

The second analysis on the SAGE data extended the exposure assessment of ambient air pollution by measuring traffic, a key source of air pollution in the San Joaquin Valley of California. Compared to simple traffic metrics, such as the distance to the nearest road, the traffic density metric incorporates traffic counts and has the advantage of accounting for multiple roads influencing a location of interest. I implemented TMLE and population intervention methods (PIM) to estimate the potential impact on term low birth weight if traffic levels were reduced in the population. This research also began exploration of controlling for contextual variables at the block-level from the U.S. census to supplement data from the birth certificates. SAGE showed the highest quartile of traffic density exposure is associated with significantly higher term low birth weight compared to the lowest quartile; however, there is not a monotonic exposure-response relation for the second and third quartiles.

In general, the studies presented in this dissertation suggest that traffic, measured as traffic-related air pollutants or traffic density, appear to contribute to adverse health when exposure occurs during pregnancy. The FACES-LITE longitudinal analysis contributed to earlier findings of a baseline effect. The SAGE analysis is a more methods-based approach, but also suggests adverse effects of prenatal exposure on birth outcomes. There is a wealth of information still to be gained from the SAGE data. Future studies will incorporate the ambient air pollution levels, measured by air pollution monitors throughout the counties established by the U.S. EPA. The birth certificate also provides additional information on

birth outcomes including preterm birth, small for gestational age, and intrauterine growth retardation. The large sample size of the SAGE dataset allows for future investigation into effect modification by county, race, socioeconomic status or other factors that may help explain the distribution of adverse birth outcomes in this population with regard to its relationship with air pollution exposure during pregnancy. Finally, additional variables from the U.S. census and more sources of data in the public record will be used to more fully characterize the study population.

In conclusion, this dissertation work included a great deal of learning beyond the scope of this paper that will be useful in future endeavors as an epidemiologist. Beginning with the challenge of obtaining institutional review board approval through the logistics of accessing computers with enough power to analyze such a large dataset with complex algorithms for model fitting procedures, I experienced the depth of effort that must be applied to such an undertaking. SAGE, in particular, included both large public data record acquisition, geo-coding the residences of the study population, and survey design and implementation for a secondary part of the study not included in this dissertation. In summary, this dissertation served as both a contribution to the literature as well as to my training as an epidemiologist and public health professional.

References

(1995). "Standardization of Spirometry, 1994 Update. American Thoracic Society." Am J Respir Crit Care Med 152(3): 1107-36.

Abbey, D. E., P. K. Mills, et al. (1991). "Long-term ambient concentrations of total suspended particulates and oxidants as related to incidence of chronic disease in California Seventh-Day Adventists." Environ Health Perspect 94: 43-50.

Adams, K. and D. Eschenbach (2004). "The genetic contribution towards preterm delivery." Semin Fetal Neonatal Med. Dec(9(6):): 445-52.

Annels, M., P. Hart, et al. (2004). "Interleukins-1,-4,-6,-10, tumor necrosis factor, transforming growth factor - b, FAS, and mannose-binding protein C gene polymorphisms in Austrailan women: Risk of preterm birth." American Journal of Obstetrics and Gynecology 191: 2056-67.

Annesi-Maesano, I., D. Moreau, et al. (2001). "In utero and perinatal complications preceding asthma." Allergy 56(6): 491-7.

Avol, E. L., W. J. Gauderman, et al. (2001). "Respiratory effects of relocating to areas of differing air pollution levels." Am J Respir Crit Care Med 164(11): 2067-72.

Baccarelli, A., R. O. Wright, et al. (2009). "Rapid DNA methylation changes after exposure to traffic particles." Am J Respir Crit Care Med 179(7): 572-8.

Barker, D. (1997). "Fetal nutrition and cardiovascular disease in later life." Br Med Bull 53: 96-108.

Barker, D., P. Gluckman, et al. (1993). " Fetal nutrition and cardiovascular disease in adult life." Lancet 341: 938-941.

Barker, D. J. (2004). "The developmental origins of adult disease." J Am Coll Nutr 23(6 Suppl): 588S-595S.

Barker, D. J. and C. Osmond (1986). "Childhood respiratory infection and adult chronic bronchitis in England and Wales." Br Med J (Clin Res Ed) 293(6557): 1271-5.

Bates, D. V. and R. Sizto (1987). "Air pollution and hospital admissions in Southern Ontario: the acid summer haze effect." Environ Res 43(2): 317-31.

Beelen, R., G. Hoek, et al. (2008). "Long-term effects of traffic-related air pollution on mortality in a Dutch cohort (NLCS-AIR study)." Environ Health Perspect 116(2): 196-202.

Behrman, R. E., Butler, A.S., eds. (2007). Preterm Birth: Causes, Consequences and Prevention. Washington, D.C., Institute of Medicine, Committee on Understanding Premature Birth and Assuring Healthy Outcomes: National Academies Press.

Bell, M. L., K. Ebisu, et al. (2007). "Ambient air pollution and low birth weight in Connecticut and Massachusetts." Environ Health Perspect 115(7): 1118-24.

Bell, M. L., Ebisu, K., Belanger, K. (2007). "The relationship between air pollution and low birth weight: effects by mother's age, infant sex, co-pollutants, and pre-term births. ." Environmental Research 3(4): 1118-1124.

Bernstein, J. A., N. Alexis, et al. (2004). "Health effects of air pollution." J Allergy Clin Immunol 114(5): 1116-23.

Bettegowda, V. R., T. Dias, et al. (2008). "The relationship between cesarean delivery and gestational age among US singleton births." Clin Perinatol 35(2): 309-23, v-vi.
Bjorksten, B. (1999). "The environmental influence on childhood asthma." Allergy 54 Suppl 49: 17-23.
Bloom, B., R. A. Cohen, et al. (2009). "Summary health statistics for U.S. children: National Health Interview Survey, 2007." Vital Health Stat 10(239): 1-80.
Bobak, M. (2000). "Outdoor air pollution, low birth weight, and prematurity." Environ Health Perspect 108(2): 173-6.
Bobak, M. and D. Leon (1999). " The effect of air pollution on infant mortality appears specific for respiratory causes in the postneonatal period." Epidemiology. 10: 666-670.
Bobak, M. and D. A. Leon (1992). "Air pollution and infant mortality in the Czech Republic, 1986-88." Lancet 340(8826): 1010-4.
Bobak, M. and D. A. Leon (1999). "Pregnancy outcomes and outdoor air pollution: an ecological study in districts of the Czech Republic 1986-8." Occup Environ Med 56(8): 539-43.
Bodnar, L., M. Davidian, et al. (2004). "Marginal Structural Models for Analyzing Causal Effects of Time-Dependent Treatments: An application in Perinatal Epidemiology." Am J Epidemiol 159: 926-934.
Boothe, V. L. and D. G. Shendell (2008). "Potential health effects associated with residential proximity to freeways and primary roads: review of scientific literature, 1999-2006." J Environ Health 70(8): 33-41, 55-6.
Braback, L. and B. Forsberg (2009). "Does traffic exhaust contribute to the development of asthma and allergic sensitization in children: findings from recent cohort studies." Environ Health 8(1): 17.
Brauer, M., C. Avila-Casado, et al. (2001). "Air pollution and retained particles in the lung." Environ Health Perspect 109(10): 1039-43.
Brauer, M., C. Lencar, et al. (2008). "A cohort study of traffic-related air pollution impacts on birth outcomes." Environ Health Perspect 116(5): 680-6.
Brauner, E. V., L. Forchhammer, et al. (2007). "Exposure to ultrafine particles from ambient air and oxidative stress-induced DNA damage." Environ Health Perspect 115(8): 1177-82.
Briggs, D. J. (2007). "The use of GIS to evaluate traffic-related pollution." Occup Environ Med 64(1): 1-2.
Brook, R. (2007). " Updated review and controversies. Is air pollution a cause of cardiovascular disease?" Rev Environ Health 22(2):115-37(Apr-Jun).
Brotman, R. M., M. A. Klebanoff, et al. (2008). "A longitudinal study of vaginal douching and bacterial vaginosis--a marginal structural modeling analysis." Am J Epidemiol 168(2): 188-96.
Brunekreef, B. and S. T. Holgate (2002). "Air pollution and health." Lancet 360(9341): 1233-42.
Brunekreef, B., N. A. Janssen, et al. (1997). "Air pollution from truck traffic and lung function in children living near motorways." Epidemiology 8(3): 298-303.

Brunekreef, B., A. W. Stewart, et al. (2009). "Self-reported truck traffic on the street of residence and symptoms of asthma and allergic disease: a global relationship in ISAAC phase 3." Environ Health Perspect 117(11): 1791-8.
Bryan, J., Z. Yu, et al. (2004). "Analysis of longitudinal marginal structural models." Biostatistics 5(3): 361-80.
Burnett, R. T., R. E. Dales, et al. (1994). "Effects of low ambient levels of ozone and sulfates on the frequency of respiratory admissions to Ontario hospitals." Environ Res 65(2): 172-94.
CADoT, C. D. o. T. (2004). "Traffic Volumes: Annual Average Daily Traffic." from http://www.dot.ca.gov/hq/traffops/saferesr/trafdata/.
CARB. (2010). "Area Designations Maps / State and National." from http://www.arb.ca.gov/desig/adm/adm.htm.
Census, U. S. B. o. t. (2000). Census of population and housing, 2000. Summary tape file 3: technical documentation. . Washington D.C., Bureau of the Census 2000.
Churg, A., M. Brauer, et al. (2003). "Chronic exposure to high levels of particulate air pollution and small airway remodeling." Environ Health Perspect 111(5): 714-8.
Ciccone, G., F. Forastiere, et al. (1998). "Road traffic and adverse respiratory effects in children. SIDRIA Collaborative Group." Occup Environ Med 55(11): 771-8.
Collins, J. J., H. S. Kasap, et al. (1971). "Environmental factors in child mortality in England and Wales." Am J Epidemiol 93(1): 10-22.
Corburn, J., J. Osleeb, et al. (2006). "Urban asthma and the neighbourhood environment in New York City." Health Place 12(2): 167-79.
Coulton, C. J., J. Korbin, et al. (2001). "Mapping residents' perceptions of neighborhood boundaries: a methodological note." Am J Community Psychol 29(2): 371-83.
Cunningham, J., D. W. Dockery, et al. (1995). "Racial differences in the association between maternal smoking during pregnancy and lung function in children." Am J Respir Crit Care Med 152(2): 565-9.
Cunningham, J., D. W. Dockery, et al. (1994). "Maternal smoking during pregnancy as a predictor of lung function in children." Am J Epidemiol 139(12): 1139-52.
Dejmek, J., S. Selevan, et al. (1999). "Fetal growth and maternal exposure to particulate matter during pregnancy." Environ Health Perspectv(107): 475-480.
Dejmek, J., I. Solansky, et al. (2000). "The impact of polycyclic aromatic hydrocarbons and fine particles on pregnancy outcome." Environ Health Perspect 108: 1159-1164.
Delfino, R. J. (2002). "Epidemiologic evidence for asthma and exposure to air toxics: linkages between occupational, indoor, and community air pollution research." Environ Health Perspect 110 Suppl 4: 573-89.
Delfino, R. J., P. J. Quintana, et al. (2004). "Association of FEV1 in asthmatic children with personal and microenvironmental exposure to airborne particulate matter." Environ Health Perspect 112(8): 932-41.
Diaz-Sanchez, D., L. Proietti, et al. (2003). "Diesel fumes and the rising prevalence of atopy: an urban legend?" Curr Allergy Asthma Rep. Mar;3(2):146-52.

Dietert, R. R., R. A. Etzel, et al. (2000). "Workshop to identify critical windows of exposure for children's health: immune and respiratory systems work group summary." Environ Health Perspect 108 Suppl 3: 483-90.
Diez Roux, A. V. (2001). "Investigating neighborhood and area effects on health." Am J Public Health 91(11): 1783-9.
Dockery, D., Pope CW (1993). "Acute respiratory effects of particulate air pollution." Annul Rev. Public Health(15:): 107-132.
Dockery, D. W., F. E. Speizer, et al. (1989). "Effects of inhalable particles on respiratory health of children." Am Rev Respir Dis 139(3): 587-94.
Dockery, D. W., J. H. Ware, et al. (1985). "Distribution of forced expiratory volume in one second and forced vital capacity in healthy, white, adult never-smokers in six U.S. cities." Am Rev Respir Dis 131(4): 511-20.
Dominici, F., R. D. Peng, et al. (2006). "Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases." Jama 295(10): 1127-34.
Dugandzic, R., L. Dodds, et al. (2006). "The association between low level exposures to ambient air pollution and term low birth weight: a retrospective cohort study." Environ Health 5: 3.
Engel, S., H. Erichsen, et al. (2005). "Risk of spontaneous preterm birth is associated with common proinflammatory cytokine polymorphisms." Epidemiology. Jul;(16(4):): 469-77.
Engel, S., A. Olshan, et al. (2005). "Risk of small-for-gestational age is associated with common anti-inflammatory cytokine polymorphisms." Epidemiology.(Jul;16(4):): 478-86.
ERP, E. R. P. (2007). Guidelines for the Diagnosis and Management of Asthma. Bethesda, MD National Institutes of Health; National Heart, Lung, and Blood Institute: 11-34.
Fierro, M. (2000, June 15, 2000). "Carbon Monoxide." from http://coep.pharmacy.arizona.edu/air/air_quality/CARBON%20MONOXIDE.pdf.
Fuchs, K. and R. Wapner (2006). "Elective cesarean section and induction and their impact on late preterm births." Clin Perinatol 33(4): 793-801; abstract viii.
Gauderman, W. J., E. Avol, et al. (2004). "The effect of air pollution on lung development from 10 to 18 years of age." N Engl J Med 351(11): 1057-67.
Gauderman, W. J., E. Avol, et al. (2005). " Associations between childhood asthma, nitrogen dioxide, and proximity to major roadways." Epidemiology. (in press).
Gauderman, W. J., G. F. Gilliland, et al. (2002). "Association between air pollution and lung function growth in southern California children: results from a second cohort." Am J Respir Crit Care Med 166(1): 76-84.
Gauderman, W. J., R. McConnell, et al. (2000). "Association between air pollution and lung function growth in southern California children." Am J Respir Crit Care Med 162(4 Pt 1): 1383-90.
Gauderman, W. J., H. Vora, et al. (2007). "Effect of exposure to traffic on lung development from 10 to 18 years of age: a cohort study." Lancet 369(9561): 571-7.

Genereux, M., N. Auger, et al. (2008). "Neighbourhood socioeconomic status, maternal education and adverse birth outcomes among mothers living near highways." J Epidemiol Community Health 62(8): 695-700.
Ghosh, R., J. Rankin, et al. (2007). "Does the effect of air pollution on pregnancy outcomes differ by gender? A systematic review." Environ Res 105(3): 400-8.
Gilliland, F., K. Berhane, et al. (2000). "Maternal smoking during pregnancy, environmental tobacco smoke exposure and childhood lung function." Thorax. Apr;55(4):271-6.
Gilliland, F., Y. Li, et al. (2001). "Effects of maternal smoking during pregnancy and environmental tobacco smoke on asthma and wheezing in children." Am J Respir Crit Care Med 163:429-36.
Gillman, M. W. (2002). "Epidemiological challenges in studying the fetal origins of adult chronic disease." Int J Epidemiol 31(2): 294-9.
Glinianaia, S., J. Rankin, et al. (2004.). "Particulate air pollution and fetal health: a systematic review of the epidemiologic evidence." Epidemiology 15: 36-45.
Gotschi, T., J. Heinrich, et al. (2008). "Long-term effects of ambient air pollution on lung function: a review." Epidemiology 19(5): 690-701.
Gruber, S., van der Laan, M.J. (2009). Targeted Maximum Likelihood Estimation: A Gentle Introduction. Working Paper 252, U.C. Berkeley Division of Biostatistic.
Ha, E. H., J. T. Lee, et al. (2003). "Infant susceptibility of mortality to air pollution in Seoul, South Korea." Pediatrics 111(2): 284-90.
Halterman, J. S., K. A. Lynch, et al. (2009). "Environmental exposures and respiratory morbidity among very low birth weight infants at 1 year of life." Arch Dis Child 94(1): 28-32.
Hamilton, B. E., A. M. Minino, et al. (2007). "Annual summary of vital statistics: 2005." Pediatrics 119(2): 345-60.
Hanrahan, J. P., I. B. Tager, et al. (1992). "The effect of maternal smoking during pregnancy on early infant lung function." Am Rev Respir Dis 145(5): 1129-35.
HEI (2010). Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects. Boston, MA, Health Effects Institute.
Hernan, M. A. (2004). "A definition of causal effect for epidemiological research." J Epidemiol Community Health 58(4): 265-71.
Hernan, M. A., S. Hernandez-Diaz, et al. (2004). "A structural approach to selection bias." Epidemiology 15(5): 615-25.
Hernan, M. A. and J. M. Robins (2006). "Estimating causal effects from epidemiological data." J Epidemiol Community Health 60(7): 578-86.
Hirsch, T., S. K. Weiland, et al. (1999). "Inner city air pollution and respiratory health and atopy in children." Eur Respir J 14(3): 669-77.
Hoek, G., B. Brunekreef, et al. (2002). "Association between mortality and indicators of traffic-related air pollution in the Netherlands: a cohort study." Lancet 360(9341): 1203-9.
Hole, D. J., G. C. Watt, et al. (1996). "Impaired lung function and mortality risk in men and women: findings from the Renfrew and Paisley prospective population study." Bmj 313(7059): 711-5; discussion 715-6.

Holgate, S. T. and M. Peters-Golden (2003). "Introduction: the anti-inflammatory role of cysteinyl leukotriene receptor antagonists in asthma." J Allergy Clin Immunol 111(1 Suppl): S1-4.
Holgate, S. T. and R. Polosa (2006). "The mechanisms, diagnosis, and management of severe asthma in adults." Lancet 368(9537): 780-93.
Holgate, S. T., Samet, J.M, Koren, H.S., Maynard, R.L., Ed. (1999). Air Pollution and Health. London, Academic Press.
Holguin, F., S. Flores, et al. (2007). "Traffic-related exposures, airway function, inflammation, and respiratory symptoms in children." Am J Respir Crit Care Med 176(12): 1236-42.
Hu, P., Reuscher, T. (2004). Summary of Travel Trends: 2001 National Household Travel Survey., U.S. Department of Transportation.
Hubbard, A. E. and M. J. Laan (2008). "Population intervention models in causal inference." Biometrika 95(1): 35-47.
Huss, K., P. L. Naumann, et al. (2001). "Asthma severity, atopic status, allergen exposure and quality of life in elderly persons." Ann Allergy Asthma Immunol 86(5): 524-30.
Huynh, M., T. J. Woodruff, et al. (2006). "Relationships between air pollution and preterm birth in California." Paediatr Perinat Epidemiol 20(6): 454-61.
Hyslop, N. P., S. G. Brown, et al. (2003). "California Regional PM10/PM2.5 Air Quality Study (CRPAQS) data quality summary reports. Final report prepared for San Joaquin Valleywide Air Pollution Study Agency and California Air Resources Board, Sacramento, CA, by Sonoma Technology, Inc., Petaluma, CA, STI-999242-2310-FR, February."
Ihorst, G., T. Frischer, et al. (2004). "Long- and medium-term ozone effects on lung growth including a broad spectrum of exposure." Eur Respir J 23(2): 292-9.
Jaakkola, J. J., P. Ahmed, et al. (2006). "Preterm delivery and asthma: a systematic review and meta-analysis." J Allergy Clin Immunol 118(4): 823-30.
Jaakkola, J. J. and M. Gissler (2007). "Are girls more susceptible to the effects of prenatal exposure to tobacco smoke on asthma?" Epidemiology 18(5): 573-6.
Jalaludin, B., T. Mannes, et al. (2007). "Impact of ambient air pollution on gestational age is modified by season in Sydney, Australia." Environ Health 6: 16.
Jedrychowski, W., F. Perera, et al. (2009). "Gender differences in fetal growth of newborns exposed prenatally to airborne fine particulate matter." Environ Res 109(4): 447-56.
Jerrett, M., A. Arain, et al. (2005). "A review and evaluation of intraurban air pollution exposure models." J Expo Anal Environ Epidemiol 15(2): 185-204.
Jerrett, M., R. T. Burnett, et al. (2005). "Spatial analysis of air pollution and mortality in Los Angeles." Epidemiology 16(6): 727-36.
Kan, H., G. Heiss, et al. (2008). "Prospective analysis of traffic exposure as a risk factor for incident coronary heart disease: the Atherosclerosis Risk in Communities (ARIC) study." Environ Health Perspect 116(11): 1463-8.
Kannan, S., D. P. Misra, et al. (2006). "Exposures to airborne particulate matter and adverse perinatal outcomes: a biologically plausible mechanistic framework for exploring potential effect modification by nutrition." Environ Health Perspect 114(11): 1636-42.

Kannan, S., D. P. Misra, et al. (2007). "Exposures to airborne particulate matter and adverse perinatal outcomes: a biologically plausible mechanistic framework for exploring potential." Cien Saude Colet 12(6): 1591-602.
Katsouyanni, K. (2003). "Ambient air pollution and health." Br Med Bull 68: 143-56.
Kelly, K. E., D. A. Wagner, et al. (2003). "Characterization of exhaust particles from military vehicles fueled with diesel, gasoline, and JP-8." J Air Waste Manag Assoc 53(3): 273-82.
Kim, J. J. (2004). "Ambient air pollution: health hazards to children." Pediatrics 114(6): 1699-707.
Kim, J. J., S. Smorodinsky, et al. (2004). "Traffic-related air pollution near busy roads: the East Bay Children's Respiratory Health Study." Am J Respir Crit Care Med 170(5): 520-6.
Ko, H., J. W. Hogan, et al. (2003). "Estimating causal treatment effects from longitudinal HIV natural history studies using marginal structural models." Biometrics 59(1): 152-62.
Kunzli, N., R. McConnell, et al. (2003). "Breathless in Los Angeles: the exhausting search for clean air." Am J Public Health 93(9): 1494-9.
Kunzli, N. and I. B. Tager (2000). "Long-term health effects of particulate and other ambient air pollution: research can progress faster if we want it to." Environ Health Perspect 108(10): 915-8.
Lacasana, M., A. Esplugues, et al. (2005). "Exposure to ambient air pollution and prenatal and early childhood health effects." Eur J Epidemiol. 20(2):183-99.
Latzin, P., M. Roosli, et al. (2009). "Air pollution during pregnancy and lung function in newborns: a birth cohort study." Eur Respir J 33(3): 594-603.
Lave, J. R. and L. B. Lave (1977). "Measuring the effectiveness of prevention: I." Milbank Mem Fund Q Health Soc 55(2): 273-89.
Law, C. (1995). "Fetal influences on adult hypertension." J Hum Hypertension 9: 649-651.
Lee, B. E., E. H. Ha, et al. (2003). "Exposure to air pollution during different gestational phases contributes to risks of low birth weight." Hum Reprod 18(3): 638-43.
Leem, J. H., B. M. Kaplan, et al. (2006). "Exposures to air pollutants during pregnancy and preterm delivery." Environ Health Perspect 114(6): 905-10.
Lefebvre, G., J. A. Delaney, et al. (2008). "Impact of mis-specification of the treatment model on estimates from a marginal structural model." Stat Med 27(18): 3629-42.
Li, Y., F. Gilliland, et al. (2000). "Effects of in utero and environmental tobacco smoke exposure on lung function in boys and girls with and without asthma." Am J Respir Crit Care Med; 162(6): 2097-104.
Lin, M., Y. Chen, et al. (2002). "The influence of ambient coarse particulate matter on asthma hospitalization in children: case-crossover and time-series analyses." Environ Health Perspect 110(6): 575-81.
Lipfert, F. W., J. Zhang, et al. (2000). "Infant mortality and air pollution: a comprehensive analysis of U.S. data for 1990." J Air Waste Manag Assoc 50(8): 1350-66.
Liu, J., M. Ballaney, et al. (2008). "Combined inhaled diesel exhaust particles and allergen exposure alter methylation of T helper genes and IgE production in vivo." Toxicol Sci 102(1): 76-81.

Liu, S., D. Krewski, et al. (2003). "Association between gaseous ambient air pollutants and adverse pregnancy outcomes in Vancouver, Canada." Environ Health Perspect 111(14): 1773-8.
London, S. J., W. James Gauderman, et al. (2001). "Family history and the risk of early-onset persistent, early-onset transient, and late-onset asthma." Epidemiology 12(5): 577-83.
Longo, L. D. and K. S. Ching (1977). "Placental diffusing capacity for carbon monoxide and oxygen in unanesthetized sheep." J Appl Physiol 43(5): 885-93.
Loomis, D., M. Castillejos, et al. (1999). " Air pollution and infant mortality in Mexico City." Epidemiology. 10: 118-123.
Maisonet, M., T. J. Bush, et al. (2001). "Relation between ambient air pollution and low birth weight in the Northeastern United States." Environ Health Perspect 109 Suppl 3: 351-6.
Maisonet, M., A. Correa, et al. (2004). "A review of the literature on the effects of ambient air pollution on fetal growth." Environ Res 95: 106-115.
Maroziene, L. and R. Grazuleviciene (2002). "Maternal exposure to low-level air pollution and pregnancy outcomes: a population-based study." Environ Health 1(1): 6.
Martin, C. R., Y. F. Brown, et al. (2009). "Nutritional practices and growth velocity in the first month of life in extremely premature infants." Pediatrics 124(2): 649-57.
McConnell, R., K. Berhane, et al. (2002). "Asthma in exercising children exposed to ozone: A cohort study." Lancet 359: p. 386-91.
McConnell, R., K. Berhane, et al. (2003). " Prospective Study of Air Pollution and Bronchitic Symptoms in Children with Asthma." Am J Respir Crit Care Med; 168(7): 790-797.
McCreanor, J., P. Cullinan, et al. (2007). "Respiratory effects of exposure to diesel traffic in persons with asthma." N Engl J Med 357(23): 2348-58.
Mead, J. (1980). "Dysanapsis in normal lungs assessed by the relationship between maximal flow, static recoil, and vital capacity." American Review of Respiratory Disease 121: 339-342.
Meng, Y. Y., M. Wilhelm, et al. (2008). "Are frequent asthma symptoms among low-income individuals related to heavy traffic near homes, vulnerabilities, or both?" Ann Epidemiol 18(5): 343-50.
Meng, Y. Y., M. Wilhelm, et al. (2007). "Traffic and outdoor air pollution levels near residences and poorly controlled asthma in adults." Ann Allergy Asthma Immunol 98(5): 455-63.
Mortimer, K., R. Neugebauer, et al. (2008). "Air pollution and pulmonary function in asthmatic children: effects of prenatal and lifetime exposures." Epidemiology 19(4): 550-7; discussion 561-2
Mortimer, K. M., A. Fallot, et al. (2003). "Evaluating the use of a portable spirometer a study of pediatric asthma." Chest 123: 1899-1907.
Mortimer, K. M., L. M. Neas, et al. (2002). "The effect of air pollution on inner-city children with asthma." Eur Respir J 19(4): 699-705.
Mortimer, K. M., R. Neugebauer, et al. (2005). "An application of model-fitting procedures for marginal structural models." Am J Epidemiol 162(4): 382-8.

Mortimer, K. M., I. B. Tager, et al. (2000). "The effect of ozone on inner-city children with asthma: identification of susceptible subgroups." Am J Respir Crit Care Med 162(5): 1838-45.
Moshammer, H., G. Hoek, et al. (2006). "Parental smoking and lung function in children: an international study." Am J Respir Crit Care Med 173(11): 1255-63.
Mudway, I. S. and F. J. Kelly (2000). "Ozone and the lung: a sensitive issue." Mol Aspects Med 21(1-2): 1-48.
Mudway, I. S., N. Stenfors, et al. (2004). "An in vitro and in vivo investigation of the effects of diesel exhaust on human airway lining fluid antioxidants." Arch Biochem Biophys 423(1): 200-12.
Neuberger, M. and H. Moshammer (2004). "[Suspended particulates and lung health]." Wien Klin Wochenschr 116 Suppl 1: 8-12.
Nukui T, Day R, et al. (2004). "Maternal/newborn GSTT1 null genotypes contributes to risk of preterm, low birthweight infants." Pharmacogenetics 14: 569-576.
Nukui, T., R. D. Day, et al. (2004). "Maternal/newborn GSTT1 null genotype contributes to risk of preterm, low birthweight infants." Pharmacogenetics 14(9): 569-76.
O'Connor, G. T., L. Neas, et al. (2008). "Acute respiratory health effects of air pollution on children with asthma in US inner cities." J Allergy Clin Immunol 121(5): 1133-1139 e1.
Oberdorster, G. and M. Utell (2002). "Ultrafine particles in the urban air: to the respiratory tract -- and beyond?" Environ Health Perspect 110: A440-A441.
Oftedal, B., W. Nystad, et al. (2009). "Long-term traffic-related exposures and asthma onset in schoolchildren in oslo, norway." Environ Health Perspect 117(5): 839-44.
OHIR. (2010). "Birth Certificates." from http://www.cdph.ca.gov/certlic/birthdeathmar/Pages/default.aspx.
Ostro, B. D., R. Broadwin, et al. (2000). "Coarse and fine particles and daily mortality in the Coachella Valley, California: a follow-up study." J Expo Anal Environ Epidemiol 10(5): 412-9.
Parker, J. D., P. Mendola, et al. (2008). "Preterm birth after the Utah Valley Steel Mill closure: a natural experiment." Epidemiology 19(6): 820-3.
Parker, J. D., T. J. Woodruff, et al. (2005). "Air pollution and birth weight among term infants in California." Pediatrics 115(1): 121-8.
Pearson, R. L., H. Wachtel, et al. (2000). "Distance-weighted traffic density in proximity to a home is a risk factor for leukemia and other childhood cancers." J Air Waste Manag Assoc 50(2): 175-80.
Peden, D. B. (2001). "Air pollution in asthma: effect of pollutants on airway inflammation." Ann Allergy Asthma Immunol 87(6 Suppl 3): 12-7.
Penfold, B., Lurmann, F., Alcorn, S. (2009). Summary of SAGE traffic metrics. K. Mortimer, Padula, A., Gale, S., Ross, Z., Jerrett, M., Sonoma Technology, Inc.
Pereira, L., D. Loomis, et al. (1998). "Association between air pollution and intrauterine mortality in Sao Paulo, Brazil." Environ Health Perspect Jun;106((6)): 325-9.
Perera, F., W. Y. Tang, et al. (2009). "Relation of DNA methylation of 5'-CpG island of ACSL3 to transplacental exposure to airborne polycyclic aromatic hydrocarbons and childhood asthma." PLoS ONE 4(2): e4488.

Peters, J. M., E. Avol, et al. (1999). "A study of twelve Southern California communities with differing levels and types of air pollution. II. Effects on pulmonary function." Am J Respir Crit Care Med 159(3): 768-75.
Petersen, M. L., Y. Wang, et al. (2006). "Assessing the effectiveness of antiretroviral adherence interventions. Using marginal structural models to replicate the findings of randomized controlled trials." J Acquir Immune Defic Syndr 43 Suppl 1: S96-S103.
Pinkerton, J. E. (2007). "Sulfur dioxide and nitrogen oxides emissions from U.S. pulp and paper mills, 1980-2005." J Air Waste Manag Assoc 57(8): 901-6.
Pinkerton, K. E. and J. P. Joad (2006). "Influence of air pollution on respiratory health during perinatal development." Clin Exp Pharmacol Physiol 33(3): 269-72.
Ponce, N., K. Hoggatt , et al. (2005). "Preterm birth: the interaction of traffic-related air pollution with economic hardship in Los Angeles neighborhoods." Am J Epidemiol. Jul 15;162(2):: 140-8. Epub 2005 Jun 22.
Ponce, N. A., K. J. Hoggatt, et al. (2005). "Preterm birth: the interaction of traffic-related air pollution with economic hardship in Los Angeles neighborhoods." Am J Epidemiol 162(2): 140-8.
Pope, C. A., 3rd, R. T. Burnett, et al. (2004). "Cardiovascular mortality and long-term exposure to particulate air pollution: epidemiological evidence of general pathophysiological pathways of disease." Circulation 109(1): 71-7.
Pope, C. A., 3rd and D. W. Dockery (2006). "Health effects of fine particulate air pollution: lines that connect." J Air Waste Manag Assoc 56(6): 709-42.
Pope, C. A., 3rd, D. W. Dockery, et al. (1991). "Respiratory health and PM10 pollution. A daily time series analysis." Am Rev Respir Dis 144(3 Pt 1): 668-74.
Pope, C. A., 3rd, M. J. Thun, et al. (1995). "Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults." Am J Respir Crit Care Med 151(3 Pt 1): 669-74.
Postlethwait, E. M., S. D. Langford, et al. (1995). "NO2 reactive absorption substrates in rat pulmonary surface lining fluids." Free Radic Biol Med 19(5): 553-63.
Pryor, W. A., G. L. Squadrito, et al. (1995). "The cascade mechanism to explain ozone toxicity: the role of lipid ozonation products." Free Radic Biol Med 19(6): 935-41.
Ramsey, C. D., J. C. Celedon, et al. (2005). "Predictors of disease severity in children with asthma in Hartford, Connecticut." Pediatr Pulmonol 39(3): 268-75.
Rasmussen, F., D. R. Taylor, et al. (2002). "Risk factors for airway remodeling in asthma manifested by a low postbronchodilator FEV1/vital capacity ratio: a longitudinal population study from childhood to adulthood." Am J Respir Crit Care Med 165(11): 1480-8.
Rasmussen, F., D. R. Taylor, et al. (2002). "Risk factors for hospital admission for asthma from childhood to young adulthood: a longitudinal population study." J Allergy Clin Immunol 110(2): 220-7.
Riedler, J., C. Braun-Fahrlander, et al. (2001). "Exposure to farming in early life and development of asthma and allergy: a cross-sectional survey." Lancet 358(9288): 1129-33.

Ritz, B. and M. Wilhelm (2008). "Ambient air pollution and adverse birth outcomes: methodologic issues in an emerging field." Basic Clin Pharmacol Toxicol 102(2): 182-90.
Ritz, B., M. Wilhelm, et al. (2007). "Ambient air pollution and preterm birth in the environment and pregnancy outcomes study at the University of California, Los Angeles." Am J Epidemiol 166(9): 1045-52.
Ritz, B., M. Wilhelm, et al. (2006). "Air pollution and infant death in southern California, 1989-2000." Pediatrics 118(2): 493-502.
Ritz, B. and F. Yu (1999). "The effect of ambient carbon monoxide on low birth weight among children born in Southern California between 1989 and 1993." Environ Health Perspect 107: 17-25.
Ritz, B. and F. Yu (1999.). "The effect of ambient carbon monoxide on low birth weight among children born in Southern California between 1989 and 1993." Environ Health Perspect 107:: 17-25.
Ritz, B., F. Yu, et al. (2000). "Effect of air pollution on preterm birth among children born in Southern California between 1989 and 1993." Epidemiology 11: 502-511.
Ritz, B., F. Yu, et al. (2002). " Ambient air pollution and risk of birth defects in Southern California." American Journal of Epidemiology 155: 17-25.
Robins, J. M., M. A. Hernan, et al. (2000). "Marginal structural models and causal inference in epidemiology." Epidemiology 11(5): 550-60.
Rogers, J. F., S. J. Thompson, et al. (2000). "Association of very low birth weight with exposures to environmental sulfur dioxide and total suspended particulates." Am J Epidemiol 151(6): 602-13.
Rojas-Martinez, R., R. Perez-Padilla, et al. (2007). "Lung function growth in children with long-term exposure to air pollutants in Mexico City." Am J Respir Crit Care Med 176(4): 377-84.
Sagiv, S. K., P. Mendola, et al. (2005). "A time-series analysis of air pollution and preterm birth in Pennsylvania, 1997-2001." Environ Health Perspect 113(5): 602-6.
Salam, M. T., J. Millstein, et al. (2005). "Birth Outcomes and Prenatal Exposure to Ozone, Carbon Monoxide, and Particulate Matter: Results from the Children's Health Study." Environ Health Perspect. 113(11): 1638-1644.
Salmond, C., P. Crampton, et al. (1999). "Asthma prevalence and deprivation: a small area analysis." J Epidemiol Community Health 53(8): 476-80.
Salvi, S. S., C. Nordenhall, et al. (2000). "Acute exposure to diesel exhaust increases IL-8 and GRO-alpha production in healthy human airways." Am J Respir Crit Care Med 161(2 Pt 1): 550-7.
Sarnat, J. A., J. Schwartz, et al. (2001). "Fine particulate air pollution and mortality in 20 U.S. cities." N Engl J Med 344(16): 1253-4.
Savitz, D. A. and L. Feingold (1989). "Association of childhood cancer with residential traffic density." Scand J Work Environ Health 15(5): 360-3.
Schlesinger, R. B., N. Kunzli, et al. (2006). "The health relevance of ambient particulate matter characteristics: coherence of toxicological and epidemiological inferences." Inhal Toxicol 18(2): 95-125.

Schwartz, J. (1989). "Lung function and chronic exposure to air pollution: a cross-sectional analysis of NHANES II." Environ Res 50(2): 309-21.
Schwartz, J. (1991). "Particulate air pollution and daily mortality: a synthesis." Public Health Rev 19(1-4): 39-60.
Schwartz, J. (2001). "Air pollution and blood markers of cardiovascular risk." Environ Health Perspect 109 Suppl 3: 405-9.
Schwartz, J. (2004). "Air Pollution and Children's Health." Pediatrics 113(4): 1037-1043.
Schwartz, J., D. Gold, et al. (1990). "Predictors of asthma and persistent wheeze in a national sample of children in the United States. Association with social class, perinatal events, and race." Am Rev Respir Dis 142(3): 555-62.
Schwartz, J. and L. M. Neas (2000). "Fine particles are more strongly associated than coarse particles with acute respiratory health effects in schoolchildren." Epidemiology 11(1): 6-10.
Selevan, S. G., C. A. Kimmel, et al. (2000). "Identifying critical windows of exposure for children's health." Environ Health Perspect 108 Suppl 3: 451-5.
Sherrill, D., C. J. Holberg, et al. (1990). "Differential rates of lung growth as measured longitudinally by pulmonary function in children and adolescents." Pediatr Pulmonol 8(3): 145-54.
Sies, H. (1991). "Oxidative stress: from basic research to clinical application." Am J Med 91(3C): 31S-38S.
Sin, D. D. and S. F. Man (2005). "Chronic obstructive pulmonary disease as a risk factor for cardiovascular morbidity and mortality." Proc Am Thorac Soc 2(1): 8-11.
Sinisi, S. and M. van der Laan (2004). "The deletion/substitution/addition algorithm in loss function based estimation: Applications in genomics." Journal of Statistical Methods in Molecular Biology 3(1).
Sinisi, S. E., R. Neugebauer, et al. (2006). "Cross-validated bagged prediction of survival." Stat Appl Genet Mol Biol 5: Article12.
Sinisi, S. E. and M. J. van der Laan (2004). "Deletion/substitution/addition algorithm in learning with applications in genomics." Stat Appl Genet Mol Biol 3: Article18.
Slama, R., V. Morgenstern, et al. (2007). "Traffic-related atmospheric pollutants levels during pregnancy and offspring's term birth weight: a study relying on a land-use regression exposure model." Environ Health Perspect 115(9): 1283-92.
Sorlie, P., W. Kannel, et al. (1989). " Mortality associated with respiratory function and symptoms in advanced age. The Framingham Study." Amer Rev Respir Di 140:379-84.
Squadrito, G. L., R. Cueto, et al. (2001). "Quinoid redox cycling as a mechanism for sustained free radical generation by inhaled airborne particulate matter." Free Radic Biol Med 31(9): 1132-8.
Steffensen, F. H., H. T. Sorensen, et al. (2000). "Low birth weight and preterm delivery as risk factors for asthma and atopic dermatitis in young adult males." Epidemiology 11(2): 185-8.
Stick, S. (2000). "Pediatric origins of adult lung disease. 1. The contribution of airway development to paediatric and adult lung disease." Thorax 55(7): 587-94.

Stillerman, K. P., D. R. Mattison, et al. (2008). "Environmental exposures and adverse pregnancy outcomes: a review of the science." Reprod Sci 15(7): 631-50.
Stocks, J. and C. Dezateux (2003). "The effect of parental smoking on lung function and development during infancy." Respirology 8(3): 266-85.
Strand, V., S. Rak, et al. (1997). "Nitrogen dioxide exposure enhances asthmatic reaction to inhaled allergen in subjects with asthma." Am J Respir Crit Care Med 155(3): 881-7.
Strickland, M. J., M. Klein, et al. (2009). "Ambient air pollution and cardiovascular malformations in Atlanta, Georgia, 1986-2003." Am J Epidemiol 169(8): 1004-14.
Studnicka, M., E. Hackl, et al. (1997). "Traffic-related NO2 and the prevalence of asthma and respiratory symptoms in seven year olds." Eur Respir J 10(10): 2275-8.
Suh, Y. J., H. Kim, et al. (2009). "Different effects of PM10 exposure on preterm birth by gestational period estimated from time-dependent survival analyses." Int Arch Occup Environ Health 82(5): 613-21.
Tager, I., J. Balmes, et al. (2005 .). "Chronic exposure to ambient ozone and lung function in young adults." Epidemiol Nov;16(6):751-9.
Tager, I., S. Weiss, et al. (1983). "Longitudinal study of the effects of maternal smoking on pulmonary function in children." New England Journal of Medicine 309: 699-703.
Tager, I. B. (2008). "The effects of second-hand and direct exposure to tobacco smoke on asthma and lung function in adolescence." Paediatr Respir Rev 9(1): 29-37; quiz 37-8.
Tager, I. B., J. Balmes, et al. (2005). "Chronic exposure to ambient ozone and lung function in young adults." Epidemiology 16(6): 751-9.
Tewari, A. and N. P. Shukla (1991). "Air pollution--adverse effects of sulfur dioxide." Rev Environ Health 9(1): 39-46.
Thurston, G. D., K. Ito, et al. (1994). "Respiratory hospital admissions and summertime haze air pollution in Toronto, Ontario: consideration of the role of acid aerosols." Environ Res 65(2): 271-90.
Tonne, C., S. Melly, et al. (2007). "A case-control analysis of exposure to traffic and acute myocardial infarction." Environ Health Perspect 115(1): 53-7.
Turnovska, T. H. and B. I. Marinov (2009). "The influence of air pollution during intrauterine development and early childhood on respiratory functions at later age." Int J Hyg Environ Health 212(5): 519-32.
U.S.EPA. (2008). "Air Trends." from www.epa.gov/air/airtrends/index.html.
U.S.EPA. (2008). "National Emissions Inventory (NEI) Air Pollutant Emissions Trends Data." from www.epa.gov/ttn/chief/trends/.
U.S.EPA. (2010). "Nitrogen Dioxide." from http://www.epa.gov/airtrends/nitrogen.html.
U.S.EPA. (2010). "Clean Air Act." from http://www.epa.gov/air/caa/.
van den Hooven, E. H., V. W. Jaddoe, et al. (2009). "Residential traffic exposure and pregnancy-related outcomes: a prospective birth cohort study." Environ Health 8: 59.
van der Laan, M. from http://www.stat.berkeley.edu/ ~laan/Software/.
van der Laan, M. J., E. C. Polley, et al. (2007). "Super learner." Stat Appl Genet Mol Biol 6: Article25.
Wang, L. Y., Y. Zhong, et al. (2005). "Direct and indirect costs of asthma in school-age children." Prev Chronic Dis 2(1): A11.

Wang, X., H. Ding, et al. (1997). " Association between air pollution and low birth weight: a community- based study." Environ Health Perspect 105:: 514-520.
Wang, X., D. W. Dockery, et al. (1993). "Pulmonary function growth velocity in children 6 to 18 years of age." Am Rev Respir Dis 148(6 Pt 1): 1502-8.
Wang, X., D. Wypij, et al. (1994). "A longitudinal study of the effects of parental smoking on pulmonary function in children 6-18 years." Am J Respir Crit Care Med 149(6): 1420-5.
Wang, X., B. Zuckerman, et al. (2002). "Maternal cigarette smoking, metabolic gene polymorphism, and infant birth weight." JAMA 287(2): 241-2.
Watson, J. G., L. W. Chen, et al. (2008). "Source apportionment: findings from the U.S. Supersites Program." J Air Waste Manag Assoc 58(2): 265-88.
Watson, J. P., P. Cowen, et al. (1996). "The relationship between asthma admission rates, routes of admission, and socioeconomic deprivation." Eur Respir J 9(10): 2087-93.
Weller, B. (2010). "Asthma and Air Pollution." 2010, from http://www.arb.ca.gov/research/asthma/asthma.htm.
Wheeler, B. W. and Y. Ben-Shlomo (2005). "Environmental equity, air quality, socioeconomic status, and respiratory health: a linkage analysis of routine data from the Health Survey for England." J Epidemiol Community Health 59(11): 948-54.
Wilcox, A. J. (1993). "Birth weight and perinatal mortality: the effect of maternal smoking." Am J Epidemiol 137(10): 1098-104.
Wilcox, A. J. and I. T. Russell (1983). "Perinatal mortality: standardizing for birthweight is biased." Am J Epidemiol 118(6): 857-64.
Wilhelm, M. and B. Ritz (2003). "Residential proximity to traffic and adverse birth outcomes in Los Angeles county, California, 1994-1996." Environ Health Perspect 111(2): 207-16.
Wilhelm, M. and B. Ritz (2005). "Local variations in CO and particulate air pollution and adverse birth outcomes in Los Angeles County, California, USA." Environ Health Perspect 113(9): 1212-21.
Wjst, M., P. Reitmeir, et al. (1993). "Road traffic and adverse effects on respiratory health in children." Bmj 307(6904): 596-600.
Wong, E., J. Gohlke, et al. (2004). "Assessing the health benefits of air pollution reduction for children." Environ Health Perspect 112(2): 226-32.
Woodruff, T., J. Grillo, et al. (1997). "The relationship between selected causes of postneonatal infant mortality and particulate air pollution in the United States." Environ Health Perspect 105: 608-612.
Woodruff, T. J., J. D. Parker, et al. (2003). "Disparities in exposure to air pollution during pregnancy." Environ Health Perspect 111(7): 942-6.
Woodruff, T. J., J. D. Parker, et al. (2006). "Fine particulate matter (PM2.5) air pollution and selected causes of postneonatal infant mortality in California." Environ Health Perspect 114(5): 786-90.
Xepapadaki, P., Y. Manios, et al. (2009). "Association of passive exposure of pregnant women to environmental tobacco smoke with asthma symptoms in children." Pediatr Allergy Immunol 20(5): 423-9.

Xu, X., H. Ding, et al. (1995). "Acute effects of total suspended particles and sulfur dioxides on preterm delivery: a community-based cohort study." Arch Environ Health 50: 407-415.
Yang, C. Y., C. C. Chang, et al. (2003). "Evidence for increased risks of preterm delivery in a population residing near a freeway in Taiwan." Arch Environ Health 58(10): 649-54.
Yang, I. A., K. M. Fong, et al. (2008). "Genetic susceptibility to the respiratory effects of air pollution." Thorax 63(6): 555-63.
Zeger, S. L., F. Dominici, et al. (1999). "Harvesting-resistant estimates of air pollution effects on mortality." Epidemiology 10(2): 171-5.
Zeger, S. L., D. Thomas, et al. (2000). "Exposure measurement error in time-series studies of air pollution: concepts and consequences." Environ Health Perspect 108(5): 419-26.
Zeka, A., S. J. Melly, et al. (2008). "The effects of socioeconomic status and indices of physical environment on reduced birth weight and preterm births in Eastern Massachusetts." Environ Health 7: 60.
Zhu, Y., W. C. Hinds, et al. (2002). "Concentration and size distribution of ultrafine particles near a major highway." J Air Waste Manag Assoc 52(9): 1032-42.
Zhu, Y., T. Kuhn, et al. (2006). "Comparison of daytime and nighttime concentration profiles and size distributions of ultrafine particles near a major highway." Environ Sci Technol 40(8): 2531-6.

List of Appendices

Appendix 1: Abbreviations (as defined in text):

AADT	Annual average daily traffic
AIC	Akaike's information criterion
ACS	American Cancer Society
ASOSMOG	Study and Adventist Study of Smog
BS	Black Smoke
CARB	California Air Resources Board
CHS	Children's Health Study
CO	Carbon Monoxide
CRPAQS	California Regional $PM_{10}/PM_{2.5}$ Air Quality Study
D/S/A	Deletion/Substitution/Addition algorithm
D/S/A	Deletion/Substitution/Addition (algorithm)
DWTD	Distance weighted traffic density
ETA	Experimental treatment assignment
FACES	Fresno Asthmatic Children's Environment Study
FACES-LITE	FACES – Lifetime Exposure
FEF_{25}	Forced expiratory flow at 25% of FVC
FEF_{25-75}	Forced expiratory flow between 25% and 75% of FVC
FEF_{50}	Forced expiratory flow at 50% of FVC
FEF_{75}	Forced expiratory flow at 75% of FVC
FEV_1	Forced expiratory flow in one second
FVC	Forced vital capacity
GINA	Global Initiative for Asthma
GIS	Geographical information system
GST	Glutathione S-transferase
IPTW	Inverse-probability of treatment weight
IUGR	Intrauterine growth retardation
LBW	Low birth weight (< 2500 g)
LMP	Last menstrual period
LUR	Land use regression
MSM	Marginal Structural Model
NAAQS	National Ambient Air Quality Standards
NHANES	National Health and Nutrition Examination Survey
NIEHS	National Institute for Environmental Health Science
NO	Nitrous oxide
NO_2	Nitrogen Dioxide
NO_x	Nitrogen oxides
O_2	Oxygen
O_3	Ozone
OR	Odds ratio
PAHs	Polycyclic aromatic hydrocarbons
PEF	Peak expiratory flow rate

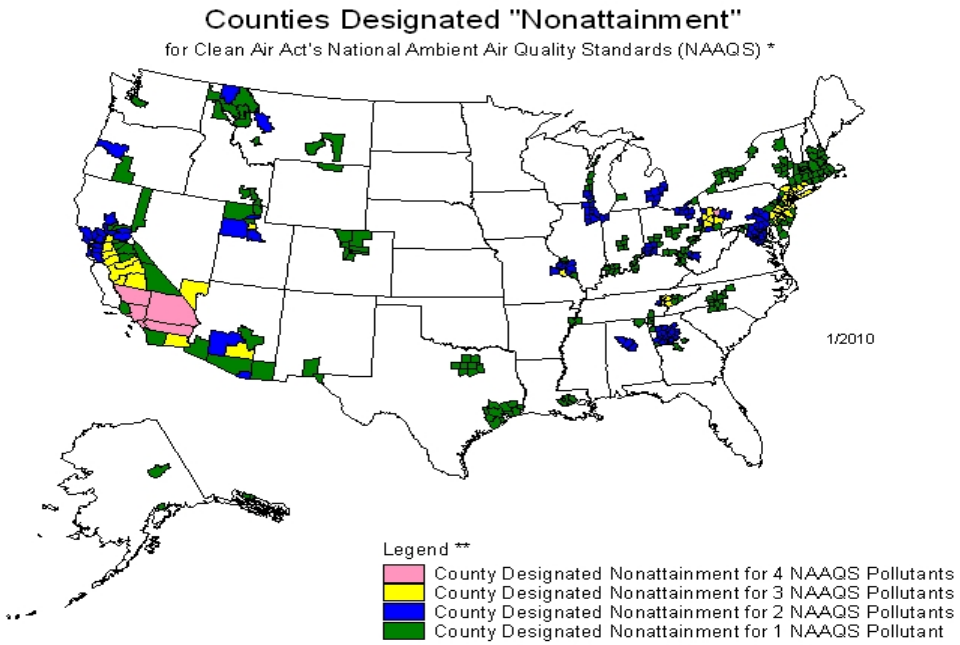
PFT	Pulmonary Function Test
PIM	Population intervention model
PM_{10}	Particulate matter less than or equal to 10 μm in aerodynamic diameter
$PM_{2.5}$	Particulate matter less than or equal to 2.5 μm in aerodynamic diameter
PTB	Preterm birth
RCTs	Randomized controlled trials
SAGE	Study of Air pollution, Genetics and the Early life events
SES	Socioeconomic status
SGA	Small for gestational age
SHS	Secondhand exposure to tobacco smoke
SIDS	Sudden infant death syndrome
SO_2	Sulfur dioxide
TMLE	Targeted Maximum Likelihood Estimation
TSP	Total suspended particles
U.S. EPA	U.S. Environmental Protection Agency
VLBW	Very low birth weight
VOCs	Volatile organic compounds
WHO	World Health Organization

Appendix 2. Ambient Air Quality Standards at the State, National and International levels as of November 2008.

Pollutant	Averaging Time	California Standards	Federal Standards	WHO Guidelines
Nitrogen dioxide (NO_2)	Annual	57 (0.030 ppm)	100 (0.053 ppm)	40
	1 hour	339 (0.18 ppm)	--	200
Carbon monoxide (CO)	8 hour	10 (9 ppm)	10 (9 ppm)	--
	1 hour	23 (20 ppm)	40 (35 ppm)	--
Respirable Particle Matter (PM_{10})	Annual	20	--	20
	24 hour	50	150	50
Fine Particle Matter ($PM_{2.5}$)	Annual	12	15	10
	24 hour	--	35	25
Ozone (O_3)	8 hour	137 (0.07 ppm)	147 (0.075 ppm)	100
	1 hour	180 (0.09 ppm)	--	--
Sulfur dioxide (SO_2)	Annual	--	80 (0.3 ppm)	--
	24 hour	105 (0.04 ppm)	365 (0.14 ppm)	20
	3 hour	--	--	--
	1 hour	655 (0.25 ppm)	--	--
	10 minute	--	--	500

Concentrations measured in $\mu g/m^3$ unless otherwise noted.
(Ref: http://www.arb.ca.gov/research/aaqs/aaqs2.pdf, www.epa.gov/air/criteria.html, http://www.who.int/mediacentre/factsheets/fs313/en/index.html)

Appendix 3. Map of counties in the U.S. designated as "nonattainment" as of January 2010.



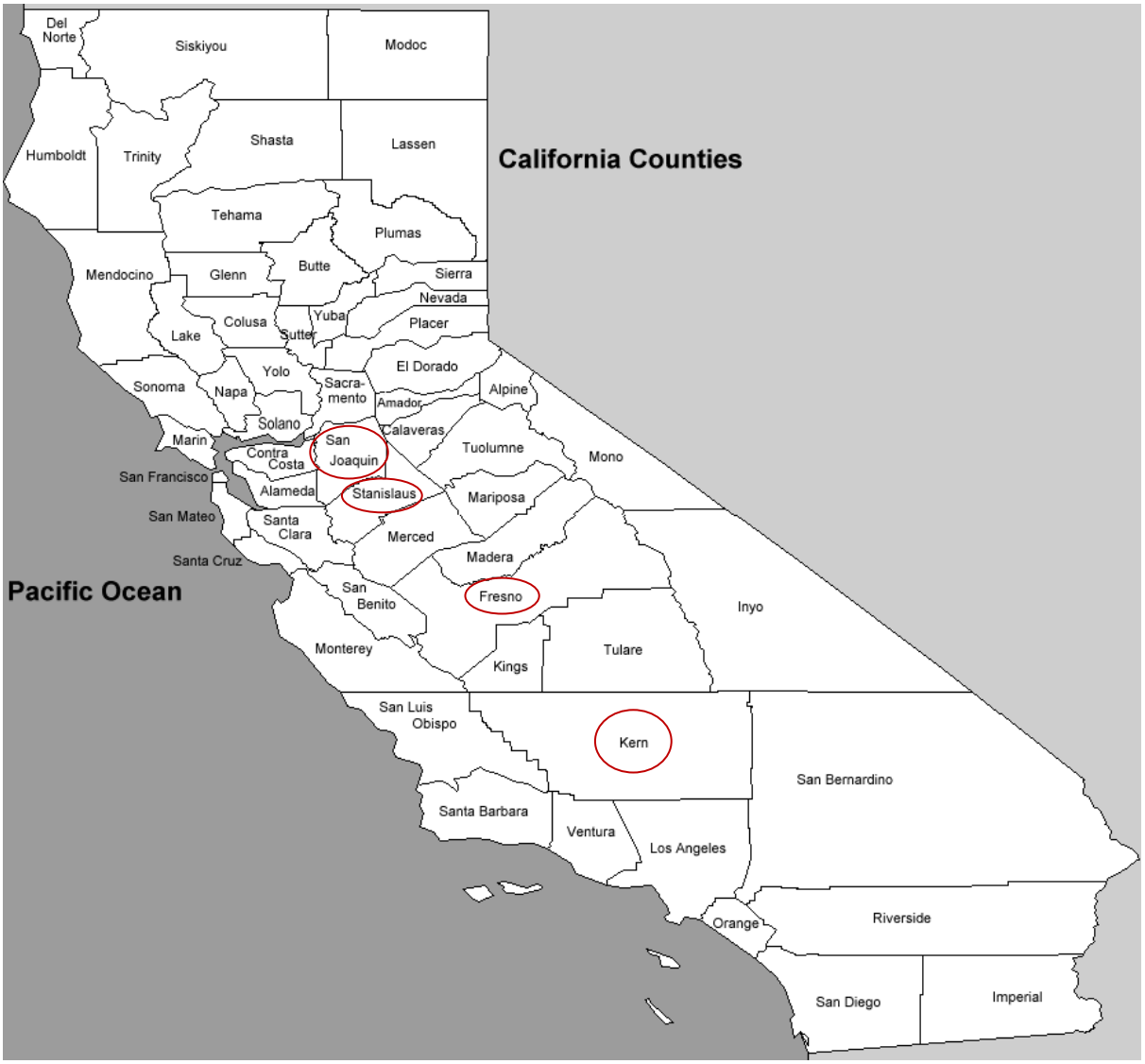
Guam - Piti and Tanguisson Counties are designated nonattainment for the SO2 NAAQS
Puerto Rico - Mun. of Guaynamo is designated nonattainment for the PM10 NAAQS

* The National Ambient Air Quality Standards are health standards for lead, carbon monoxide, sulfur dioxide, ground level 8-hr ozone, and particulate matter (PM-10 and PM2.5). There are no nitrogen dioxide nonattainment areas.

** Partial counties, those with part of the county designated nonattainment and part attainment, are shown as full counties on the map.

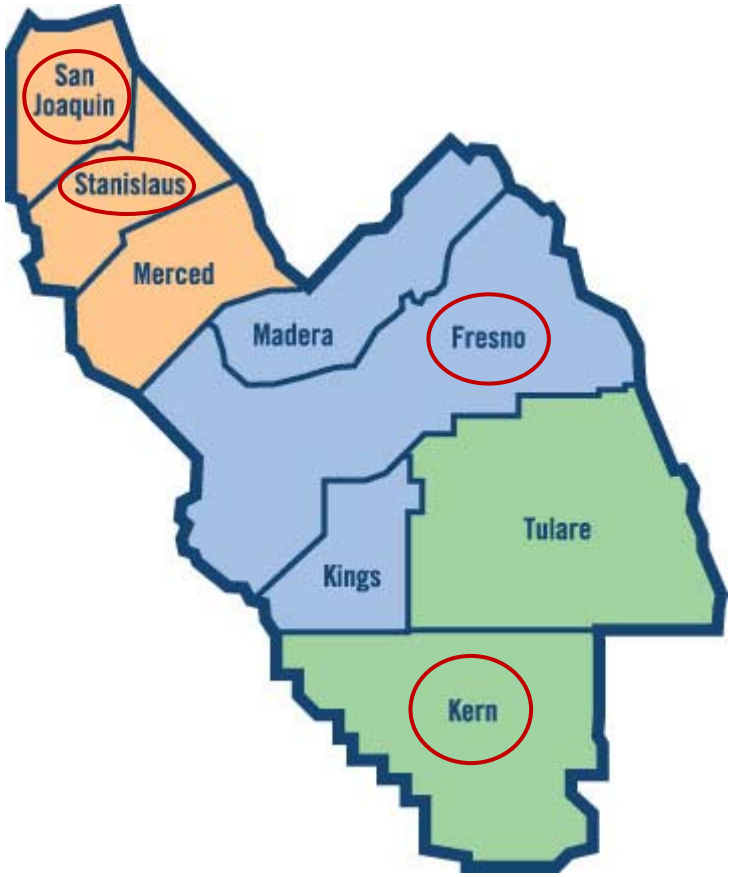
(Ref: http://www.epa.gov/oar/oaqps/greenbk/mapnpoll.html)

Appendix 4. Map of California counties (red circled counties are included in SAGE study population).



http://www.all-birth-records.com/california_counties.html

Appendix 5: Map of counties in San Joaquin Valley of California (red circled counties are included in SAGE study population).



http://www.valleyair.org/general_info/aboutdist.htm