The Impact of Neighborhood Traffic Density and Deprivation on Lung Function Among Children with Asthma

By

Sara Lynn Gale

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 Ira B. Tager, Chair
Professor Alan E. Hubbard
Professor Michael L. Jerrett
Professor John D. Radke

Spring 2012

Abstract

The Impact of Neighborhood Traffic Density and Deprivation on Lung Function Among Children with Asthma
by

Sara Lynn Gale
Doctor of Philosophy in Epidemiology

University of California, Berkeley
Professor Ira B. Tager, Chair

To investigate the extent to which traffic exposure affects the lung function of children with asthma and how local neighborhood factors may modify this relation, a merge of epidemiologic, environmental health and geographic methods is necessary. People and places are linked; therefore, it is necessary to consider place-effects on health as well as environmental exposures. The Fresno Asthmatic Children's Environment Study (FACES) is a longitudinal cohort study of children with asthma in Fresno, California that followed participants from 2000-2008 to explore short-term and long-term effects of ambient air pollution on lung function (as measured by spirometry, wheeze, and asthma symptoms). With publicly available data on traffic counts in Fresno, CA from 2000-2008, I built a spatial model of traffic exposure that varies both temporally and spatially for the FACES cohort. To capture and quantify neighborhood characteristics, I constructed individual neighborhoods based on global positioning software (GPS) data and walking distances around participant homes. To evaluate neighborhood deprivation, I collected geographic information system (GIS) data on parks, grocery stores, bus stops, etc. from publicly available sources and created an index based on Item Response Theory. To assess the marginal risk difference of lung function among children with asthma exposed to high levels of traffic pollution and those exposed to lower levels of traffic pollution (as measured by traffic density), I apply semi-parametric causal inference methods and use Targeted Maximum Likelihood Estimation (TMLE). More FACES participants who live in high deprivation neighborhoods are also farther away from high traffic areas. Neighborhood deprivation, as defined by a combination of GIS variables in this study, does not track well with US Census poverty. The marginal change in lung function from exposure to high neighborhood traffic to lower neighborhood traffic, without stratification for neighborhood deprivation, is -0.233 (95% CI -0.338, -0.129). The results can be interpreted as—the average decrease of FEV_1 is 0.233 L, or there is a 12% reduction in lung function. Either neighborhood deprivation does not modify the effect of traffic on lung function or there is not enough data to evaluate this type of effect modification. The findings indicate that neighborhood exposure to traffic adversely affects lung function among the FACES cohort of children with asthma.

Table of Contents

Acknowledgements

Many thanks to Boriana Pratt for data management and programming, and the FACES team of investigators and field staff. I thank Bryan Penfold, Siena Alcorn, and Fred Lurmann at Sonoma Technology Inc. for traffic link programming and GPS data management.

Chapter 1: Introduction

Asthma is the most important disease of childhood in terms of occurrence and morbidity.[1,2] Nationally, one in twelve children (9.6%) is estimated to have asthma.[3] Billions of dollars are spent annually on the management and care of children with asthma. Asthma is the one of the leading causes of emergency department visits among children under 15 years old in the US.[4] Defined as the chronic inflammation of the pulmonary airways, asthma is a condition where the airways become inflamed, produce mucous, and undergo episodic, reversible narrowing that results in wheezing and shortness of breath. Over the long-term, these processes lead to remodeling of the airways characterize by chronic inflammatory cells, increased airway wall thickness and decreased airway diameters—especially in airways <2mm in diameter. The net result of these chronic processes is reduced lung function. Among asthmatics, there are many airway irritants or exposures that can trigger an attack: allergens, air pollution, tobacco smoke, mold, endotoxin, viral respiratory infections, stress, and exercise. In females asthma may worsen during menstrual cycles and pregnancy. While the proximate cause and cure of asthma remain unknown, and we have considerable understanding of the factors associated with asthma prevalence and exacerbations, many unanswered questions remain about specific triggers and individual differences in the asthma phenotype.

Hippocrates states in an aphorism, "Of diseases and ages, certain of them are well- or ill-adapted to different seasons, places . . ." Asthma is reflective of this aphorism. In urban communities throughout the United States, it is estimated that up to one in five children has asthma. The burden of disease is not shared equally. Prevalence is the highest among poor children (13.5%) and non-Hispanic black children (17%).[3] Neighborhoods characterized by a higher percentage of minorities, lower incomes, inadequate housing, and ambient air pollution are correlated positively with asthma hospitalization rates.[5] Neighborhood characteristics such as violence, poverty, and poor housing have been associated with asthma prevalence and morbidity.[6]

The general health of populations and its relation to neighborhood factors is a topic that is challenging to study because of the variability in the methodology and data available: the choice of neighborhood characteristics, the quantitative evaluation of those characteristics, definition of neighborhood boundaries, the statistical methods to analyze the characteristics, and the pathway or rationale for the relation or non-relation. Neighborhood characteristics can fit into two categories: contextual and integral. Contextual factors are those social factors that describe a neighborhood, such as unemployment and poverty and are derived from summaries of individual-level data. Integral factors are the physical factors found in the neighborhood environment, like a park or an auto mechanic shop and are independent of individual-level characteristics. Both types of neighborhood factors are linked to the full panoply of health outcomes—with more research emphasis on factors that compose a deprived neighborhood instead of an advantaged one. In addition to the aforementioned challenges of this research topic, one must also consider individual-level factors to complete the picture of a neighborhood.[7]

It is well established that residents of deprived areas exhibit poorer health compared to those from more affluent areas.[8-10] Neighborhood characteristics are related to a variety of health outcomes—mortality, behavior, mental health, stress, low birth weight, obesity, asthma, and cardiovascular disease.[9,11-15] In a 30 year retrospective analysis on housing and health trends with the National Health and Nutrition Examination Survey (NHANES), researchers found positive trends with poor housing stock and Body Mass Index (BMI), blood lead levels, high blood pressure and cardiovascular disease, as well as increased asthma prevalence with more

broken and barred windows.[16] While methods may differ, researchers have consistently found that health outcome variation is related to area-level characteristics, independent of individual factors.[17] Cardiovascular disease has been shown to have strong associations with neighborhood factors in some studies, but researchers argue that the pathway to disease is through neighborhood social interaction with an individual's environment.[18] Adverse birth outcomes are often left unexplained by individual-level factors;[19] therefore, more studies are examining neighborhood-level factors. These aggregated factors include residential segregation,[20] crime,[21,22] economic disadvantage,[21-23] and low education;[23,24] and they have all been linked to adverse birth outcomes, such as preterm birth and low birth weight. In a recent study on reported crime and perception of crime, neighborhood crime perception was the factor related most to having an infant who was small for gestational age.[25] Perceived neighborhood crime and disorder often result in fear and can explain adverse health conditions. [26]

Asthma shares many of the same associations with neighborhood characteristics as most other health outcomes. Increased asthma prevalence and asthma hospitalizations have been associated with levels of deprivation in New Zealand and England.[27,28] Poor housing stock, overcrowding, homelessness, and air quality are related to emergency department visits for difficulty breathing.[29] Communities with more violence are associated with an increased asthma risk.[30] A number of studies have reported associations between community violence and increased childhood asthma diagnoses,[31] exacerbations[32,33] and hospitalizations.[34] The exposure to violence (controlling for socioeconomic status, housing, stress and negative life events) has been linked to greater asthma morbidity,[35] but the direct pathway is unknown.[36] Like cardiovascular disease, researchers have found a relationship between asthma morbidity and stress.[37,38] Stress exacerbates asthma symptoms and causes greater bronchoconstriction in asthmatic patients,[39] and, when compared to healthy controls, asthma patients exposed to stress have been shown to have increased production of cytokines implicated in the pathophysiology of asthma and its exacerbation.[39] Asthma patients with lower socioeconomic status (SES) are noted to have higher chronic stress and perceived threat.[40]

Asthma symptoms and exacerbations are good indicators of the burden and management of the disease, but lung function (forced expiratory volume (FEV) and forced vital capacity (FVC) measurements as measured by spirometry) among asthmatics is a well-studied to assess substantial airflow impairment and long-term mortality. Forced expiratory volume in one second (FEV_1) has been shown to be the most closely related airflow measure to chronic respiratory symptoms.[41] FEV_1 and ratios of FEV_1 and FVC are strong predictors of all cause mortality, not just respiratory, and cardiovascular disease.[42,43] There is a large body of literature about lung function metrics, and while FEV_1 has been used most frequently, FEF_{25-75} (forced expiratory flow between 25% and 75% of vital capacity) more closely reflects the site of important airway changes that are related to obstruction.[44]

There are many approaches to measure spatial deprivation and advantage at the local level. Often, GIS data from disparate sources are aggregated from the start; researchers must "stitch" together datasets for different times and study areas.[45] Indicators of neighborhood deprivation in the US most often include social and physical factors obtained from the decadal US Census, e.g., education, employment, family composition, housing stock, household income, poverty, and residential stability.[46] Looking beyond the limits of the census data, other research has evaluated data on crime, retail availability, services, physical disorder, social capital and reciprocity.[26,47] Determinants of deprivation reported in the literature include items such as

economic disadvantage, unemployment, education, household characteristics, and housing conditions.[12] In addition, crime, Toxic Release Inventory (TRI) sites, and alcohol outlets are being used as markers of contextual and integral neighborhood disadvantage. For example, violent crimes and property crimes are both associated with deprivation (as defined by income inequality) and low social capital in the US; the authors of this work dubbed crime a "mirror for the social environment."[48] Access to alcohol outlets is another key component of area-level deprivation. Researchers in New Zealand found proximity to alcohol outlets to be strongly associated with social deprivation.[49] Researchers in Oakland, CA identified a clustering of TRI sites in West Oakland, a historically poor, deprived area.[50] Neighborhood quality can also be measured by factors such as walkability, which can be calculated based on land use, street connectivity and residential density.[51] Neighborhood air pollution, traffic and noxious land uses can be included as integral factors used to describe a small area environment.[51,52] While much of the research on how social factors relate to health uses neighborhood-level exposures, little work has been done, in the context of studies of health interventions, to identify which factors should be measured, how to best classify neighborhoods as disadvantaged or advantaged and how to validate the measurements.

Along with the heterogeneity in the definition of deprivation or advantage and no generally agreed upon standard to quantify neighborhood characteristics, there are problems with the definition and measurement of neighborhood units. While many studies have reported an association between area-level characteristics and health outcomes, in general, neighborhoods have yet to be defined more accurately than a census or geopolitical boundary. For some residents, a census geographic unit may not capture adequately important details of their neighborhood.[53] Boundaries defined to serve geopolitical ends (e.g., taxation, government services and statistics) often are not well-suited to identify specific characteristics of neighborhoods that are linked to health.[54] Neighborhood perception can vary by a person's age, gender, social status, and physical geographic characteristics.[53,55,56] There is limited research that measures a neighborhood in terms other than census geographic units, despite reports that smaller neighborhood units have been shown to provide a more meaningful and exact estimate of area effects.[54] Since many neighborhood factors are an aggregate of individual-level attributes, some researchers argue that neighborhood factors based on aggregation are not identifiable and are over emphasized in health research.[7] While neighborhood research is a new, promising area for understanding adverse health outcomes, most of the studies I cite for a relation between health and neighborhood factors use data aggregated to the census or other government defined boundary level.

Once the data collection of deprivation and neighborhood advantage variables is complete and the neighborhood boundaries are set, researchers then must quantify the neighborhood experience. As with any data analysis, there are various approaches, and with neighborhood analyses these approaches fall into two categories: one performed outside of the GIS in a statistical model or within the GIS as a spatial model. A method often used to analyze neighborhood data is a traditional hierarchical or multi-level model. This method often is plagued by the curse of dimensionality—or sparse cells—and non-identifiability, which means that there is little to no overlap between covariate distributions of the exposed and unexposed.[7,57] Neighborhood factors can be used like any other independent variable in a statistical model—i.e., binary or continuous counts, rates, or percentages. The variables can be combined to create a composite of deprivation with principle components analysis (PCA), factor analysis, propensity

scores, or item response theory (IRT) to name a few. Spatial analyses often examine clustering and autocorrelation, or more plainly, the degree to which things that are closer in space are more alike. Cluster analyses are an example of the autocorrelation of related points that can display disease patterns as well as exposure patterns and can facilitate identification of "hot spots." There are myriad choices to examine and quantify neighborhood characteristics and there is no gold standard or perfect approach to understand the data.

Part of the neighborhood experience, but beyond strict physical and social attributes, involves exposure assessment of local sources of pollution, containments, and toxins. Among the most ubiquitous anthropogenic exposures in the US, is traffic-generated air pollution. Air pollution in the US, like other countries, was largely a result of the burning of soft coal until the middle of the twentieth century. Deadly smog episodes in Donora, Pennsylvania;[58] London, England;[59] and Meuse Valley, Belgium[60] provided the need for air pollution regulation.[61] The Clean Air Act of 1963 setup the Environmental Protection Agency and the underpinnings of the regulatory framework that surround air pollution legislation in the US today. While background levels of ambient air pollution have declined drastically since the switch to hard coal, more resources were dedicated to the health research of air pollution during the latter part of the twentieth century. Hence, more epidemiologic evidence began to mount by the 1960s--exposure to air pollution was linked to adverse health outcomes.[62-64] Regulation in the US was and is currently based on levels of single pollutants under the Clean Air Act, and, while it makes theoretical sense to place restrictions on individual constituents of air pollution, the ambient environment constitutes a complex mixture of air pollution that represent the "true" exposure. More scientists are calling for research that investigates mixtures of air pollution in multipollutant models or targeting mixtures of emissions from a source.[65] Moreover, exposure to air pollution can be extremely variable—in the US air pollution varies by coast, season, region, and among and within cities. Exposure also varies due to individual and neighborhood deprivation. It is necessary to take these complex attributes into account to estimate an exposure that is relevant. To measure single pollutants and ignore spatial and temporal variation is not an appropriate or germane depiction of the environment in which we are exposed.

Air pollution has been linked to a variety of health metrics, but the strongest evidence for a pathway from exposure to disease exists with cardiovascular disease,[66,67] asthma exacerbations [68] and all cause mortality.[69,70] Furthermore, traffic-related air pollution is linked to respiratory illnesses,[70] but the causal evidence is viewed as more suggestive than determinate.[71] Ambient air pollution consistently has been found to trigger asthma symptoms and to be associated with asthma prevalence, but air pollution exposure as the cause of asthma onset is not strongly supported. There are a few studies that have show new onset asthma cases in relation to NO_2[72] and ozone exposure,[73] like the Children's Health Study; however, more evidence with a more detailed air pollution exposure assessment and a more rigorous asthma criterion are needed. Studies have shown that there is excess asthma morbidity associated with ambient air pollution,[74,75] and there is a stronger relation with traffic-related pollution and asthma morbidity.[75-77] More research is emerging about early life exposures to ambient air pollution and the development of asthma, with the recent work of Clark *et al*. on a population based cohort of children revealing that traffic-related exposures result in the highest odds for an asthma diagnosis by the age of 3 or 4 in children.[78]

An estimated 27% of greenhouse gases is attributed to mobile source emissions in the US and, of that percentage, passenger cars, trucks and heavy duty vehicles account for 81% of the

pollution caused by this sector.[79] The brunt of traffic pollution occurs when vehicles are cold-started, slowed or idling. Measures of traffic often are used as a proxy for mixtures of ambient air pollution, because it is easier to measure the source instead of one pollutant at a time or the complex interaction and aging of pollutants occurring in the environment. It is challenging, if not impossible, to distinguish the individual contributions of pollutants, because the effects are not independent; climate, topology, and sources all interact and vary based on place. The measurement quality of traffic ranges from crude estimates to refined spatio-temporal models, partly because of the ease of computing, but mostly due to the availability and accessibility of detailed traffic count data. In turn, different metrics of traffic can result in different estimates of exposures and effects.[80] More studies employ crude traffic metrics such as residential distance to roadways, type of vehicle (diesel/unleaded gasoline), roadway designations (freeway versus side street) within a residential buffer, self-reported traffic density and traffic counts near a residence obtained from a public agency.[71] However, distances to roadways and vehicle count alone do not capture the variability in traffic exposure at the local level.[71] Density metrics through time and space with meteorology data can better estimate traffic exposure than proximity measures to different road way classes or vehicle routes (bus or truck lines).[24] Interpolation methods can predict traffic or pollution over a defined space with spatial averaging, nearest monitor, inverse distance weighting and kriging.[24] Other prediction models such as land-use regression can incorporate more independent variables such as population density, land-use, elevation, and meteorology to predict traffic through space.[24]

Ultimately, the best estimate of traffic-related air pollution exposure for an individual may be a hybrid model that incorporates both personal monitoring and real-time traffic counts.[24] Researchers in Brooklyn, NY built a small area model to understand traffic-related air pollution in a densely populated urban environment with elevated and below grade major sources of traffic (bridge and highway).[81] This study used personal backpack monitors, local meteorology, continuous traffic counters, vehicle class and speed to get a refined characterization of the spatial traffic-related pollution. Exposure to traffic-related air pollution was dependent on time, space, meteorology, and topology—the ultrafine particle measurements changed 15-20% within the first 100m from the roadway--[81] thus, this study indicates that small area/neighborhood measurements are vital for accurate exposure classification.

Recently, environmental justice or equity has developed a quantitative base in public health. Environmental justice calls for "the fair treatment of people of all races, income, and cultures with respect to the development, implementation and enforcement of environmental laws regulations, and policies, and their meaningful involvement in the decision-making processes of the government" (Christine Todd Whitman, USEPA, 2001). Instances of environmental injustice, when a certain social group is disproportionately impacted by harmful land uses, are not new phenomena, but both public health researchers and environmental advocates have started to highlight the trend to argue for more environmentally conscious policies. Specifically, in the United States, low-income communities and people of color experience unequal exposures to lead,[82] diesel exhaust,[83] traffic emissions,[84] and other pollutants.[85] Poor people are more likely to reside closer to major roadways and point source polluters.[83]

When Congress first passed the Clean Air Act in 1963, there were provisions for setting air quality standards to protect public health; but it was not until significant expansions in 1970 amendment that "sensitive groups" were included in its language. Prior to the CAA amendments

of 1970, environmental injustices were evident among disempowered social groups; but during the civil rights movement in US, cases of environmental injustice began to be linked to health. For instance, in 1967, an eight-year old girl drowned in a garbage dump, which was sited between an elementary school and a city park in a predominately African American neighborhood. Students at nearby Texas Southern University protested the death and questioned the placement of the dump.[86] After a series of communities fought to expose environmental injustices in the 1970s and 1980s, there were calls for protective legislation to help clean up polluted areas and empower disenfranchised groups. An infamous case in North Carolina in 1982 exposed the decision to place solid waste sites to dump polychlorinated biphenyls (PCBs) in predominately minority and poor neighborhoods—based on the available scientific evidence at that time, these areas were shown to not be among the most suitable choices.[87] For years, the stretch of the Mississippi River from Baton Rouge to New Orleans was called "Cancer Alley" because of the elevated cancer prevalence in small neighborhoods in close proximity to the petrochemical corridor along the river. It was not until 1993 that the state of Louisiana issued a report that stated the health and environment of this low-income, African American community were disproportionately affected by industrial pollution.[88] In 1994, President William Clinton established Executive Order 12898 and called for each Federal Agency to make "achieving environmental justice part of its mission by identifying and addressing, as appropriate, disproportionately high and adverse human health or environmental effects."[89] This order mandated that the EPA organize an environmental justice working group, which comprises heads of relevant Federal agencies. The working group is responsible to determine what criteria must be met to achieve environmental justice and how best to plan for it. In 2011, while environmental justice may be addressed regionally or by states, the Government Accountability Office (GAO) finds that the EPA at the Federal level has not developed a clear strategy, implementation plan, or performances measures when it comes to defining environmental justice and ensuring that plans for justice are made and evaluated.[90] The Federal government as well as academic researchers does not have a gold standard to approach, quantify, and rectify environmental injustice.

From a GIS and quantitative perspective, environmental justice can be studied by comparison of distributions in time over various spaces—source exposures and subgroups. Cluster analyses, comparisons between bounded groups, simple proximity measures, or complex statistical modeling can all be applied to understand if subgroups are disproportionately located and exposed to harmful sources of pollution. However, whether the subgroups or the toxic exposure came first in time, often cannot be established. Without temporal ordering, a case for environmental justice is weakened. Based on Ripley's K statistic, researchers in Oakland, CA identified a clustering of Toxic Release Inventory (TRI) sites in West Oakland, a historically poor—a tool that compares a given distribution of points to a homogeneous Poisson distribution that is characterized by complete spatial randomness.[50] (As with other environmental justice studies in Public Health, there is no mention of whether the neighborhood existed before the TRI sites or *vice versa*.) An example of comparisons made via boundaries is a clever index of inequity by Stuart *et al.* who quantify the degree to which certain population groups live near air pollution monitoring sites. [91] Mathematically, the estimate of inequity is expressed as $Fi = \log(Zi/Ti)$, where F is the inequity experienced by a subgroup *i* and Z is the fraction of the total population of the subgroup near the pollution source (defined by a distance or buffer around the source); and T is the fraction of the total population of the subgroup within the larger boundary,

like a county. Thus, F is positive when there is inequity and a larger percent of the subgroup live within areas of high pollution. F will be negative when a larger percent of the subgroup live outside of areas of high pollution. There needs to be further unbiased quantitative analyses of environmental justice like those listed above and investigations of temporality—did the subgroup move to the harmful site; was the harmful site placed where the subgroup is; and were there other options available? The analyses above lack temporal ordering, but is that necessary to shape policies on human health and land use? Do we need to know what was present first to make policy changes that do not allow people to live in close proximity to harmful environmental exposures? Or is it okay, as long as a subgroup is not overrepresented?

For epidemiologists and public health researchers, environmental justice brings to light the need to evaluate our environment more carefully, to study population health inequities, and to examine how the environment and social factors act together to cause adverse health outcomes. The main reason why epidemiologists have a difficult time studying environmental justice and its relation to health outcomes is because we have not characterized precisely the environments under study—in terms of exposure assessment. When measuring air pollution, we often lack adequate air monitoring data that is relevant to population locations and is of sufficient temporal and spatial resolution for use in health effects assessments, particularly in urban environments in which spatial heterogeneity can be substantial. Time-activity data are often missing or not collected, but are necessary to understand where people spend their time to accurately assign air pollution exposure. Particularly relevant for environmental justice research, are accurate aggregated social data to define and quantify how area-level factors affect exposures and outcomes, and precise area-level boundaries to characterize what constitutes a neighborhood.

Neighborhood, defined as a discrete geographic unit, is a subjective and evolving construct that remains ill-defined. Historically, neighborhoods emerged in American cities with the decline of the walking city and the development of new modes of transportation in the mid nineteenth century.[92] Neighborhood boundaries were developed for geopolitical reasons—to contain areas for taxation, voting, government services, statistics, etc. Originally, census tracts were designed to capture local areas within a city that were similar in population characteristics and contained about 3,000 – 6,000 residents.[93]

Furthermore, data and relationships between data can be influenced by the size and/or shape of the units from which the data are reported.[94] This issue, known as the "modifiable areal unit problem" (MAUP)[95,96] is critical in the use of spatial analysis to inform community level policies and interventions. The identification of an adequate neighborhood area should be considered vis-à-vis the disciplinary framework in which the issue is approached. For example, the American Institute of Architects Architectural Graphic Standards, as well as city and regional planners continue to use the neighborhood definition developed by Perry in the 1920s[97] as a planning unit with a five-minute walking radius.[92] The MAUP has no definitive solution but remains a challenge to spatial research that should not be overlooked. Researchers need to consider both the appropriate scale and aggregation options when using spatial boundaries. (In my previous work, I choose a neighborhood based on walking distances to bus stops in Oakland because our survey sampling took place in public schools.) My previous analysis illustrates the MAUP in Oakland, CA—as the neighborhoods are changed from the census defined areas to the neighborhoods we derived with walking distances, I saw significant differences in our exposures (crime rate, alcohol outlet rate, etc.) and in general, the differences wane as the buffer sizes decrease.[98] This is expected and should highlight the need for researchers to carefully consider

neighborhood definition and ask: what is the most appropriate size and shape of neighborhoods for this research? Although much of the literature on neighborhoods and health has shown associations between the built environment, social factors, and health, scant attention has focused on neighborhood definition.

Rationale

I propose to use a model that incorporates traffic metrics through space and time with surrogate air monitoring data from the US EPA central site and home sampling. This method would provide temporal and spatial varying data at the local, neighborhood level. While much research reports an association between area-level characteristics and health outcomes in general, I need to define neighborhoods more precisely than a census or geopolitical boundary. Moreover, epidemiologic methods are used rarely to infer causality from area-level associations and explore the extent of area-level effects. While much of epidemiology relies on census designated neighborhoods,[9] my method to define a neighborhood by local area time-activity will use the individual to define his or her neighborhood by where he or she spends time.

A merge of epidemiologic, environmental health and geographic methods is necessary to understand adequately how neighborhood factors influence health. People and places are linked; therefore, it is necessary to consider place-effects on health as well as environmental exposures. People are not randomly distributed in space or time, and it is naive to treat them as such and to ignore a key role in disease processes: human interaction with the environment and other humans. "Randomization by cluster accompanied by an analysis appropriate to randomization by individual is an exercise in self-deception." [99] Much research in air pollution and social disparities has been focused on larger scale differences, such as exposures at the census tract and city levels--rather than smaller spatial scales, such as the neighborhood level. We know that air pollution and social factors can be quite variable at the city and even census tract level; to ignore the heterogeneity in these areas could lead to severe exposure misclassification.

Last, to my knowledge, there are no analyses of air pollution and lung function that apply causal methods and use a population risk difference estimate to describe the impact of exposure to air pollution on lung function. A risk difference approach has advantages over the relative risk, a misunderstood measure that epidemiologists have commonly used since Cornfield's seminal paper in 1959.[100,101] The relative risk, a ratio of the risk of disease in the exposed and unexposed, can obscure a confounder's impact in the exposed and unexposed groups and the baseline disease risk in a population. The risk difference is more applicable to continuous outcome measures, because it gives an absolute measure of disease or alteration of process. The risk difference is a better measure to assess the effect of exposure on disease at a population level, because it can be directly applied to calculate total cases of disease caused by the exposure.[102] A risk difference, an absolute change in disease frequency, is also the appropriate measure to use when estimating causal effects under the counterfactual framework. The counterfactual is unobservable, but it is the hypothetical outcome under the alternative exposure scenario. The counterfactual framework can help conceptualize exchangeability in a dataset.[103] In this study, I will use approaches to better approximate the neighborhood, traffic exposure, and the result of traffic on lung function.

Specific Aims

1. To use publicly available data on traffic counts in Fresno, CA from 2000-2008 to build a spatial model of traffic exposure and assign individual exposures that vary both temporally and spatially.
2. To quantify more precisely individual neighborhoods with global positioning software and self-reported health diaries that record time-activity; and to evaluate neighborhood deprivation with contextual and integral neighborhood factors collected from publicly available GIS data.
3. To assess the marginal risk difference of lung function among children with asthma exposed to high levels of traffic pollution, conditional on neighborhood deprivation, and those exposed to low levels of traffic pollution.

Chapter 2: Methods

Study Population

To investigate the effects of traffic exposure and neighborhood deprivation on the lung function of asthmatic children, area-level and individual-level covariates must be measured and the associations estimated in an exposure model. The data for my dissertation research come from the Fresno Asthmatic Children's Environment Study (FACES), a longitudinal cohort study of children with asthma in Fresno, California. This study collected data to explore short-term and long-term effects of ambient air pollution on lung function (as measured by spirometry, wheeze, and asthma symptoms). To be eligible for the study, participants must have lived within a 20 kilometer radius of the EPA central monitoring site in Fresno, CA at the start of the study for at least three months, be 6-11 years of age, have a physician diagnosis of asthma, and active asthma. FACES followed children (n=315) from 2000 to 2008 and assessed exposure over time to a variety of ambient and indoor air pollutants and bioaerosols including allergens, polycyclic aromatic hydrocarbons (PAHs), environmental tobacco smoke, particulate matter, and criteria gaseous air pollutants (NO_x, SO_2, CO, and Ozone and metals as well as endotoxin).[104-108] Each participant performed twice-daily spirometry and answered symptom questions in up to three 14-day panels per year. During these panel days, air pollution data were captured from the EPA Super site located in Fresno, CA. Subjects also were observed semi-annually or annually for more detailed health and laboratory evaluations.[104-108]

To supplement the central site estimates, FACES collected spatial-temporal ambient pollution data around Fresno via mobile trailers and 83 participant homes from February 2002-February 2003. In addition to these refined exposure measurements that vary in space and time, FACES collected data on the study participants' physical and social environment. Family income, household features, indoor/outdoor environment, family demographics, and smoking habits are a few factors that the study addressed. In addition to the social and physical environment characteristics, FACES collected data on how study participants move through space. Time-activity global positioning system (GPS) data for our study population was obtained from subjects during the period June 2006-September 2007. Students were randomly selected by panel days to complete a five-day GPS study. Each subject wore a GPS unit with a data-logger (Wintec WBT-100 GeoLogger, GeoStat, Atlanta, GA) and an accelerometer (MTI Actigraph, Fort Watson Beach, FL**)** with the data-logger while he or she was active. The GPS unit accuracy was <3 meters with circular error probability and 6 meters with 2D RMS (twice the distance of sigma or 95% of the time the actual position is within 6 meters of the indicated position). To complement the GPS data, subjects also kept a diary and recorded the type of activities they performed while wearing the GPS device. The GPS sampling was performed over five consecutive days--Wednesday through Sunday or Thursday through Monday—to capture weekday and weekend changes in spatial locations and activity patterns from October 2006 to October 2007. There were 134 children selected to do the GPS study. Students were studied both in the summer and during the school year to account for possible seasonal differences in outdoor activity patterns.

Study Design

This sub-study, like FACES, is a longitudinal, prospective cohort study. There are repeated measures of traffic exposure, neighborhood deprivation, and lung function during the panel days of FACES study period, 2000-2008, but the statistical analysis is a point treatment.

Methods

Traffic Exposure Assessment

I collected traffic count data for the FACES study area in Fresno County, CA from the California Department of Transportation, Fresno Council of Governments, the Fresno County Road Maintenance Division, the City of Fresno, and the City of Clovis for 2000-2008. The California post-mile marker freeway network from Bigham, *et al*. (2009) was used to geocode the CalTrans freeway data.[109] All of the other traffic data from the Fresno agencies, which consisted of major and minor collector roads and local roads, were geocoded to the ESRI US Streets Premium (October 2010). The traffic count data from the Fresno agencies are represented as 24-hour average daily counts, and the data from CalTrans are annual average daily counts. Each traffic count point was then assigned to the road network line segments. There is not a standard method to impute counts to links, but the counts were assigned based on roadway name, connectivity, and distance. The traffic count volumes were given to roadway classes based on feature class codes (FCCs), where classes 4, 3, 2 and 1 were assigned volumes for connected links up to 5, 7, 10 and 10 km from the original traffic count locations. Only connected streets receive a value; if a street ends and begins at a different location, the traffic count value does not carry though to the new location.

Next, I developed a surface layer for traffic density (measured by annual vehicles per area) based on a smoothed function kernel density that changes over time and space. This traffic model, like the model Jerrett, *et al*. applied in a longitudinal cohort study to estimate the association between traffic exposure and obesity, has the advantage over a measure like distance to major roadway; because it uses multiple roads and the vehicle count density to explain exposure.[110] Under this method, I use a kernel estimate in ArcGIS Spatial Analyst based on my traffic data points for each year and a decay function based on the distance from the roadways.[111] The kernel density function is based on the quadratic kernel function[112] described below in Equation 1, where x is the set of data points.

Equation 1.

$$K_2(\mathbf{x}) = \begin{cases} 3\pi^{-1}(1-\mathbf{x}^T\mathbf{x})^2 & \text{if } \mathbf{x}^T\mathbf{x} < 1 \\ 0 & \text{otherwise} \end{cases} \qquad (4.5)$$

The fall-off distance for the decay function is set to 500m.[113] The zone of influence for traffic generated air pollution from a roadway is debated. Some researchers have shown that $PM_{0.1}$ or ultrafine particles cannot be differentiated from background concentrations at distances greater than 300m from the roadway.[114] Other work has reported the spatial extent for vehicular pollution to reach background levels is 100-400m for elemental carbon, 200-500m for nitrogen dioxide and 100-300m for ultrafine particles.[115] However, even larger zones have been measured in Los Angeles, CA. Hu et al found that the zone of influence of traffic generated pollution can extend 1200m downwind of the freeway, and ultrafine particle concentration was found not reach background levels until about 2600m downwind of a freeway before dawn.[116] For upwind concentrations, the same group of researchers found that ultrafine particles were elevated above background levels 600m from the freeway.[116] Traffic emissions are the major source of ultrafine particles in the urban environment.[114,117] There is likely a regional component of ultrafine particles that extends beyond the decay function that I have set at 500m, but I want to focus on the immediate neighborhood traffic contribution.[118]

Neighborhood Assignment

Historically, neighborhoods emerged in American cities with the decline of the walking city and the development of new modes of transportation in the mid nineteenth century.[92] With the absence of intracity transportation, most people in the U. S. could not segregate themselves as well as they do now—by social, economic, religious, or political views.[92] Official neighborhood boundaries were developed for geopolitical reasons—to contain areas for taxation, voting, government services, statistics, etc. Originally, census tracts were designed to capture local areas within a city that were similar in population characteristics and contained about 3,000 – 6,000 residents.[93] However, a boundary does not always dictate where people spend their time or define what people view as their neighborhood. Neighborhood perception can vary by a person's age, gender, social status, and physical, geographic characteristics.[53,55,56] There is limited research that measures a neighborhood in terms other than census geographic units, despite reports that smaller neighborhood units have been shown to provide a more meaningful and accurate estimate of area effects.[54] Thus, there is no gold standard in terms of definition and assessment for research in which neighborhood is a central or important construct. Therefore, my approach is to use GPS data to define neighborhoods in the context of where participants spend their time. The location of the child's home within the urban setting plays an important role in the level of exposure that each child has to air pollution; however, where children spend their time is paramount to measure appropriate exposure. Air pollution studies often focus mainly on exposure assignment with residence data and time-activity data sometimes used to get a more accurate exposure measurement. I propose to use time-activity data to create neighborhood boundaries that are relevant to where participants spend their time near their residence and to quantify that space.

The GPS data consist of minute-by- minute location points for each study participant. To capture the immediate neighborhood near the participant's home and to minimize points associated with vehicular travel, I limited the GPS points to those within a mile of each child's residence. I am more interested in a child's local surroundings where a family could travel by foot or bike. Each FACES participant's neighborhood boundary is defined based on his or her time-activity data at a maximum of one mile from the home. With these data, I calculated the geometric mean of each participant's point cloud and the standard distance (Equation 2) between points around the geometric mean.[119] The standard distance is like an average Euclidean distance.

Equation 2.

The Standard Distance is given as:

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n} + \frac{\sum_{i=1}^{n}(y_i - \bar{Y})^2}{n}} \qquad (1)$$

where x_i and y_i are the coordinates for feature i, $\{\bar{X}, \bar{Y}\}$ represents the Mean Center for the features, and n is equal to the total number of features.

I then built circular neighborhoods with the radius set as the standard distance and the centroid of the circle as the geometric mean of the GPS points for each study participant. The centroid of the neighborhood is not constrained to the residence. For the FACES study participants for whom we did not collect GPS data, I designated their neighborhoods as the ½ mile from their residence. Therefore, the participants without GPS data have their residences as the centroid of their circular neighborhoods and a 10-15 minute walk (1/2 mile) as their radii. The choice of a ½ mile walk is a surrogate for a participant's immediate surroundings. In the US, it has been shown that people are willing to walk different distances based on their purpose—be it commuting to a bus stop or rail station, or running errands and shopping.[120] However, in general, a maximum walking distance of ¼ to ½ mile is reasonable to consider how far pedestrians are willing to walk.[121]

Each neighborhood is individual to each participant. Thus, no two children in the study will have the same neighborhood, unless they reside in the same home and did not complete the GPS study. Aggregated neighborhood data is often plagued by the curse of dimensionality—or sparse cells—and non-identifiability, which means that there is little to no overlap between covariate distributions of the exposed and unexposed.[43,51] This neighborhood method partly solves this problem with the construction of personal neighborhoods. The neighborhood boundaries for participants do change over time, if the participants move within Fresno County and stay in the FACES cohort between 2000-2008.

Neighborhood Deprivation Assessment

With the derived neighborhood boundaries, I use an area-weighted, average method to measure the characteristics of each neighborhood. I developed this method to incorporate the aggregated US Census data into individual neighborhoods.[98] Thus, the neighborhood boundaries can capture both contextual and integral factors. Contextual factors are often aggregated social factors derived from individual characteristics. Integral factors are those that are physical features of the environment. Together, these two types of neighborhood characteristics can be used to describe neighborhood deprivation. As detailed in the Introduction, deprivation is the composite state of a neighborhood that describes its lack of healthy environmental influences—there is no generally agreed upon standard measurement for deprivation.

Deprivation measures for FACES consist of demographic data from the 2000 US Census, alcohol outlet data from the California Alcohol Beverage and Control, and point source pollutant data from CaSIL (California Spatial Information Library) and the EPA's Toxic Release Inventory (TRI). I have also collected public data on access to grocery stores, establishments with cigarette licenses, parks, daycare facilities, schools, hospitals and bus stops (see Table 1 for sources and details); these datasets will help me define more accurately the social and physical environments where study participants live.

Table 1. GIS data collected for Fresno County

Dataset	Data type	Description	Source	Date Obtained	Reference Date or Date last updated
Alcohol Outlets	GIS point locations	On-sale (e.g., bars and restaurants) and off-sale (e.g., liquor stores and convenience marts) alcohol outlets currently licensed in Fresno County. Year of original license included in dataset.	California Department of Alcohol Beverage and Control http://www.abc.ca.gov/	August 19, 2010	August 19, 2010
Bus Lines and Bus Stops	GIS line file and point locations	Current Fresno City public bus routes FAX (Fresno Area Express)	City of Fresno Information Services Department http://www.fresno.gov/Government/DepartmentDirectory/InformationServices/GIS/Layers.htm	November 27, 2010	February 11, 2009
Cigarette Permits	GIS point locations	Business names and addresses for all current cigarette permits for California and past permits back to 2004 for cigarette taxes. Year of original permit included in dataset.	California State Board of Equalization	December 16, 2010	November 4, 2010
EPA Geospatial Pollution data	GIS point locations and polygons	Current locations for sites on the Superfund National Priorities List (NPL), EPA and State Treatment, Storage, Disposal facilities, Toxic Release Inventory System sites, etc. for California	US Environmental Protection Agency http://www.epa.gov/enviro/geo_data.html	November 30, 2010	October 28, 2010
Farmer's Markets	GIS point locations	Addresses of current farmer's markets for Fresno, Kern, Stanislaus, and San Joaquin counties	California Federation of Certified Farmers' Markets www.cafarmersmarkets.com	November 30, 2010	November 30, 2010
Food Stores	GIS point locations	Addresses and business information for current grocery stores, convenience stores, markets, and other types of food stores classified by the standard industrial classification (SIC) codes for Fresno, Kern, Stanislaus, and San Joaquin counties. Year of	Dun and Bradstreet www.dnb.com	November 23, 2010	November 23, 2010

		establishment included in dataset.			
Daycares	GIS point locations	Fresno County daycare provider locations	Fresno County Public Works and Planning	November 11, 2010	November 2, 2010
Hospitals	GIS point locations	Fresno County hospital locations	Fresno County Public Works and Planning	November 11, 2010	November 2, 2010
Land Parcels	GIS polygons	Current lot sizes and tax information	Fresno County Public Works and Planning	November 11, 2010	November 2, 2010
Land Use	GIS polygons	Areas zoned for particular types of use such as residential, agricultural, industrial, etc. as of Fall 2010.	Fresno County Public Works and Planning	September 30, 2010	September 9, 2010
Neighborhood Improvement Areas	GIS polygons	The City of Fresno identified older and poorer neighborhoods in 2005 that need infrastructure improvements in the "No neighborhood left behind" project.	City of Fresno Information Services Department http://www.fresno.gov/Government/DepartmentDirectory/InformationServices/GIS/Layers.htm	October 4, 2010	June 17, 2005
Parks	GIS polygons	Current public park perimeters in Fresno County	Fresno County Public Works and Planning	September 30, 2010	August 27, 2010
Schools	GIS point locations	Current public schools in Fresno County	Fresno County Public Works and Planning	November 11, 2010	November 2, 2010
Truck Routes	GIS line file	City of Fresno designated truck routes from 2005	City of Fresno Public Works Traffic Engineering	August 17, 2009	September 25, 2005
Traffic Density	GIS point locations of vehicles per 24 hours	Intersections where vehicles were counted in Fresno County 2000-2008	City of Fresno, City of Clovis, Fresno County and Fresno Council of Government	March 10, 2009 – December 2010	December 31, 2008

2000 US Census Demographics	GIS polygons	Variables expressed as a percentage of the blockgroup that include: males in management or professional occupations, households with more than 1 person per room, households with 1999 income below federal poverty level, female headed households with dependents, households on public assistance, households earning less than $30k per year, individuals with less than a high school education, and males and females unemployed from the 2000 census	US Census http://factfinder.census.gov	August 6, 2009	April 1, 2000

To understand the complex neighborhood environment, I use item response theory (IRT) to help characterize the variables that make up a deprived neighborhood with an item response model (IRM). In an IRM, several variables (in this case alcohol outlets, schools, parks, etc.) are used to make a scale that describes a latent variable. The motivation for such a model is that one does not measure directly the "latent" characteristic of interest, such as intelligence or ability as IRT is used in education research. In our case, we use this model as a convenient way to create a score of neighborhood deprivation, using the idea behind IRT as intuition, but not deluding ourselves that this is the true underlying model. Thus, we treat neighborhood deprivation as a "latent" variable, but we do so because we do not have one variable that captures deprivation—there are several. There are many characteristics for which the IRM will provide a convenient way of creating a deprivation score as we describe below. There are also many procedures for estimating IRM's, but, in general, data-adaptive procedures are used based on minimizing an estimate of the mean-squared error of prediction. I use a two parameter logistic model (2PL) and maximum likelihood estimation for this analysis. The 2PL model for the IRT of a latent variable, θ_i, for study subject *g* in the *i*th item is as follows:

Equation 3.

$$P_g(\theta_i) = \frac{1}{1 + \exp[-1.7 \cdot a_g \cdot (\theta_i - b_g)]}$$

where b_g is the difficulty parameter, which measures the probability of positive response (or a correct answer). The discrimination parameter, a_g, measures the ability of the item to distinguish between low/high levels of the latent variable. Thus, the maximum likelihood estimation, with an M x N response matrix X (with M rows and N columns) is—

Equation 4.

$$L(a, b|X) = \prod_{i=1}^{M} \prod_{g=1}^{N} P_g(\theta_i ; a_g , b_g)^{Xig} \cdot [1 - P_g(\theta_i ; a_g , b_g)]^{1-Xig}$$

For the purposes of this analysis, we had no survey or exam, so the item parameters are all of the geographic covariates, which I converted to be positive and negative (0=negative neighborhood influence, 1=positive neighborhood influence). For example, if there were an alcohol outlet in a participant's neighborhood, his or her "response" would be 0 for that item. If there were a park in that neighborhood, the response would be 1. Scoring is based on the proportion of respondents answering the item "correctly," where 0 is wrong and 1 is right. Item parameters are fixed, and we maximize the likelihood for theta or our deprivation score. Ultimately, I will use the item response score to address whether there is a causal association between the exposure to traffic and outcome lung function.

Environmental Justice Calculation

To measure environmental justice, I use a quantitative subgroup index of inequity.[91] The index measures the degree to which residents of a subgroup, say low SES or racial minority, live

near or experience sources of integral neighborhood-level pollution (e.g. distance to major roadways, TRI sites, traffic density). In my analysis, I use low income (study participants whose annual family income is less than $15k) and neighborhood deprivation to define the subgroup. Mathematically, the estimate of inequity is expressed as

Equation 5.

$$F_i = \log(Z_i/T_i)$$

where F is the inequity experienced by a subgroup *i* and Z is the fraction of the total subgroup (e.g. low-income) FACES population near high traffic pollution and T is the fraction of the total subgroup population within the FACES cohort. Thus, F is positive when there is inequity and a larger percent of the subgroup live within areas of high traffic density. F will be negative when a larger percent of the subgroup live outside of areas of high traffic density. This estimate will show children in low income households in comparison to the whole spatial boundary of Fresno. The estimate only quantifies exposures and does not assume that unhealthy outcomes occur based on any disproportions displayed by estimate.

To implement this measure of environmental justice with the individual neighborhood estimates, I will use each child's income status (with low income defined as a family annual income of less than $15k) and neighborhood deprivation with *i* as a marker for the low income sub-group. The areas of traffic-related high pollution will be defined by the neighborhood kernel density traffic volumes for Fresno County. This estimate of inequity will show if low income children or children who live in areas of high neighborhood deprivation in FACES are disproportionately located near high traffic density in comparison to the whole spatial boundary of FACES study area and entire FACES study population.

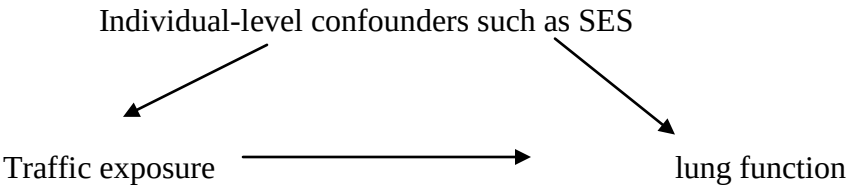
Statistical Model for Causal Inference

After developing an IRM score for the deprivation variables and assessing any inequity between deprivation and exposure to traffic, I will employ the causal inference framework to determine if lung function is affected by traffic exposure. Developing a causal estimate for traffic density exposure and lung function of a susceptible population is relevant, because traffic exposure can be changed through regulation and public awareness. Defining parameters of interest based on causal inference associations can be motivated by the concept of counterfactuals and estimation of marginal, population-level effects.[122] In this paper, I am interested in the marginal effect of high traffic exposure on lung function.
My research question is—does exposure to neighborhood traffic reduce lung function? There are three causal analyses. First, for the ½ mile walking distance neighborhoods, I will assess average lung function and its relation to low/high traffic neighborhoods. For the second and third analyses, I will assess the same question about lung function and traffic, but these estimates will be stratified by neighborhood deprivation.

There are several assumptions I have to make about the temporal relation between social disadvantage and neighborhood deprivation. Based on lack of temporality between neighborhood deprivation and individual level confounders (often, we cannot separate which came first, neighborhood deprivation factors (e.g. liquor stores) or high traffic), I consider neighborhood deprivation as an effect modifier in all analyses. The Directed Acyclic Graph

(DAG) in Figure 1 shows the relation between confounders and the outcome and exposure. Neighborhood deprivation is not displayed on the DAG in Figure 1 because I am assuming it is an effect modifier and will stratify the analysis to account for it.

Figure 1. Directed Acyclic Graph (DAG)



For this analysis, I assume that neighborhood deprivation—both integral and contextual factors—causes a different relation between traffic exposure and lung function. In general, this assumption says that the physical environment and social factors change the role that traffic exposure has on lung function. The principle hypothesis that I want to test is if traffic decreases lung function among children with asthma conditional on neighborhood deprivation.
While the FACES study is a longitudinal cohort design, this sub-study follows an observational point treatment structure, where there is one outcome and one exposure per study participant. (When there is more than one residence per participant, I can assess the health effect. However, we cannot look at a change over time; because we do not have multiple address data on all study participants.) I assess the average traffic exposure and average lung function annually for the FACES study population. The risk difference will be the change in mean lung function when exposed to high and low traffic. A risk difference is a better measure to assess the effect of exposure on disease at a population level, as it can be directly applied to calculate total cases of disease caused by the exposure.[102] This approach has advantages over the relative risk, a misunderstood measure that epidemiologists have commonly used since Cornfield's seminal paper in 1959.[100,101] The relative risk, a ratio of the risk of disease in the exposed and unexposed, can obscure a confounder's impact in the exposed and unexposed groups and the baseline disease risk in a population. I want an estimate on the scale of the probability of the outcome, because it is better for assessment of the population-level impact of changing exposures. The statistical estimator that is the most relevant to answer my research question in this study to approximate the risk difference is the estimator from a semi–parametric Targeted Maximum Likelihood Estimation (TMLE). My approach is semi-parametric, because the method makes very few assumptions about the data-generating distribution and thus, avoids specification of a fully parametric model; we will use a loss-based estimation approach.[123]

TMLE uses maximum likelihood estimation of the parameter to reduce bias in the estimate by regression of the outcome on a function of the exposure and confounders (a nuisance parameter), with an offset for the targeted estimate of the density.[124,125] The equation for the estimator of TMLE is

Equation 6.

$$\psi_n^{TMLE} = \frac{1}{n}\sum_{i=1}^{n} Q_n^*(1, W_i) - Q_n^*(0, W_i)$$

where n is the total observations, Q^*_n is the function of the outcome and confounders, W, under exposure or treatment set to 1 and Q^*_n is the function of the outcome and confounders, W, under the exposure or treatment set to 0.[125] The Q function estimation relies on the estimation nuisance parameter g(A,W) from the exposure or treatment, A, and the confounders. The tmleLite package in R (available in R 2.14.1) provides the application for TMLE in this analysis. Any model selection procedure can be used to estimate the Q portion of the likelihood. The tmleLite package employs the DSA (Deletion/Substition/Addition) algorithm (available in R 2.14.1 statistical software as an add-on package "DSA"), a data adaptive procedure, to model the relation between exposure to individual-level covariates from the FACES baseline interview, neighborhood-level deprivation, and exposure to traffic. The DSA is a data-driven learning algorithm that uses the L-2 loss function to search for a model in the entire model space of polynomials that best fits the user supplied specifications (model form, degree of interaction terms, maximum variable numbers, and powers).[126] The DSA has an "ID" option that groups the selection of observations by ID, if there are repeated measures for study subjects. While the DSA is a data adaptive procedure, I supply it with known confounders of the association between exposure to traffic and lung function (FEV_1) and use it to model the unknowable data generating distribution of these confounders, as implied by the DAG in Figure 1.

The DSA, while it is a data driven algorithm, evaluates the variables in terms of the causal hypothesis in so far as the variables are encoded in the causal graph. The variables that the DSA is allowed to consider are determined by the causal hypothesis and the researcher to get at an accurate probability distribution.

There is no causal inference without causal assumptions and hypotheses.[122] While causal inference is dependent on a set of assumptions, these assumptions can apply to traditional models as well, but none of the assumptions can be tested. First, I assume that counterfactual outcomes exist and can be estimated from the observed data. I also assume temporality in my data—the exposure precedes the outcome and the confounders precede both the outcome and the exposure. I also assume that there is no unmeasured confounding in the data with respect to the exposure and censoring. Both the exposure and the missing data, I assume to be assigned at random and missing at random, beyond the confounders that we have measured and accounted for in the models. In addition to the assumptions listed above, the data also rely on the Experimental Treatment Assignment, which states that no set of variables can deterministically restrict or assign the exposure. The ETA can be checked with a plot of the predicted probabilities for treatment over the log odds of treatment for the selected g model. The final inference and interpretation of the data relies on this set of assumptions. My analysis will help us understand if neighborhood deprivation affects exposure to traffic and lung function among children with asthma.

Chapter 3: Results

Traffic Exposure Assessment

The traffic data collected from 2000-2008 provided a snapshot of the traffic processes in Fresno County, CA. Table 2 shows the description of the data collected by agency and the resulting data that were geocoded to the street network. Many of the traffic data points were added together or divided to fit on the road network (n=7,311); therefore, the total number of points was consolidated to 5,924 for the study period. The reason for these decisions was based on the fact that for the majority of data points, we had data on both traffic directions but only one line segment, instead of two in the road network. For example, if we had both eastbound and westbound traffic and there was only one line segment in the road network depicting both directions, the eastbound and westbound points were added together. If there was one point (labeled "Band" or "both") and two road segments, the data point was halved.

Table 2.

Data Source	Years	Total Raw Data Points	Total Geocoded Data Points(%)	Total fit to Street Network by Road type
City of Fresno	2000-2008	4,102	4,012 (98%)	-
City of Clovis	2002-2008	623	623 (100%)	-
Fresno County	2000-2002, 2004-2008	6,955	6,740 (97%)	-
CalTrans	2000-2008	1,992	1,860 (93%)	-
Total	2000-2008	13,672	13,235 (97%)	5,924 (43%)

The traffic data reveal adequate coverage for the County of Fresno with data gaps in more rural areas (Figures 2a and b).

Figure 2a. Map of Fresno County, CA traffic data points (n=5924) overlaid on the County street network 2000-2008 with the FACES participants' residences (n=518)

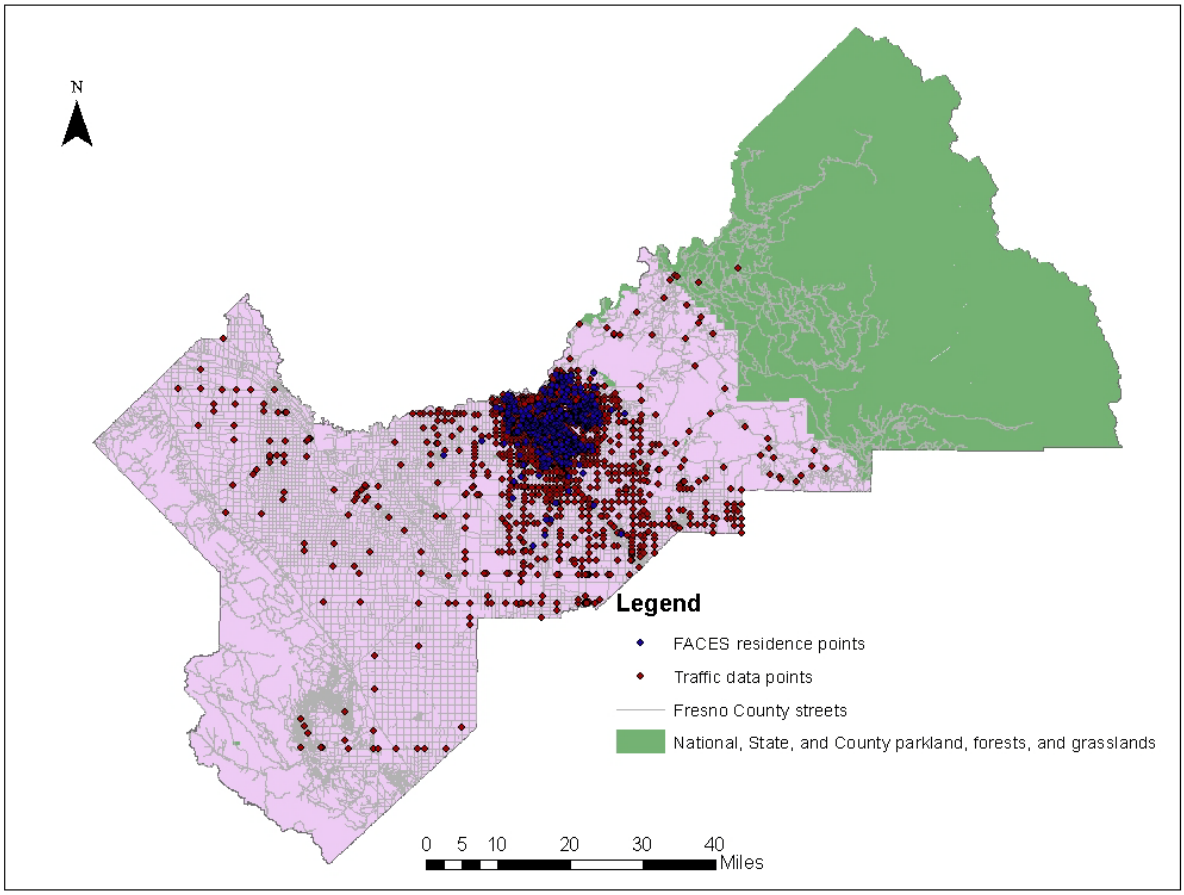
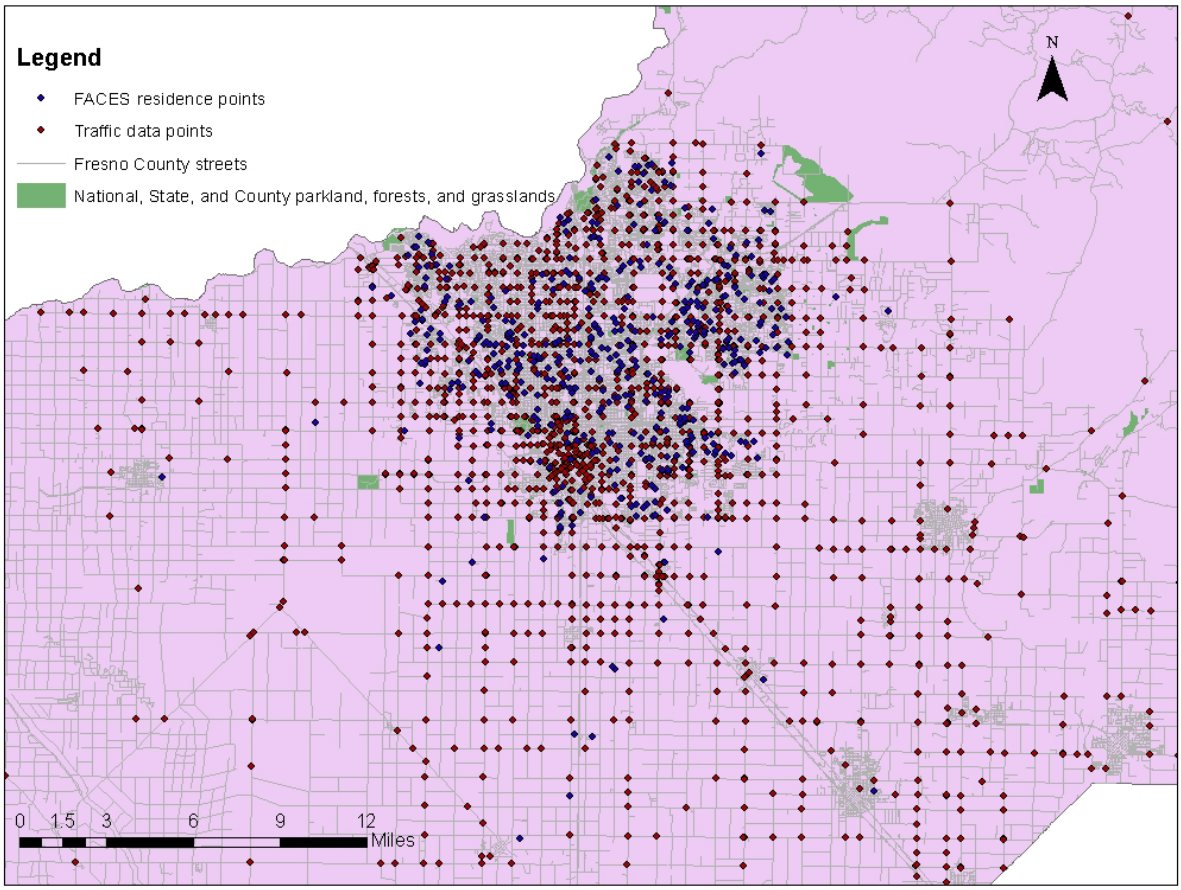


Figure 2b. Zoomed in map of Fresno County, CA FACES study area with traffic data points (n=5924) overlaid on the County street network 2000-2008 and the FACES participants' residences (n=518)



The traffic volume data points were then linked to the street network, such that each line segment in the street network is assigned a volume based on the nearest input. The maximum distance for a volume assignment from a data point was set to 5, 7, 10 and 10 kilometers for road classes 4, 3, 2, and 1 respectively. The roadway classes are coarsely defined: 4=local neighborhood road, 3=secondary and connecting road, 2=primary road, and 1= highway. Figure 3 shows the street network covered with the traffic volumes.

Figure 3a. Map of Fresno County, CA traffic volume assignment to street network 2000-2008

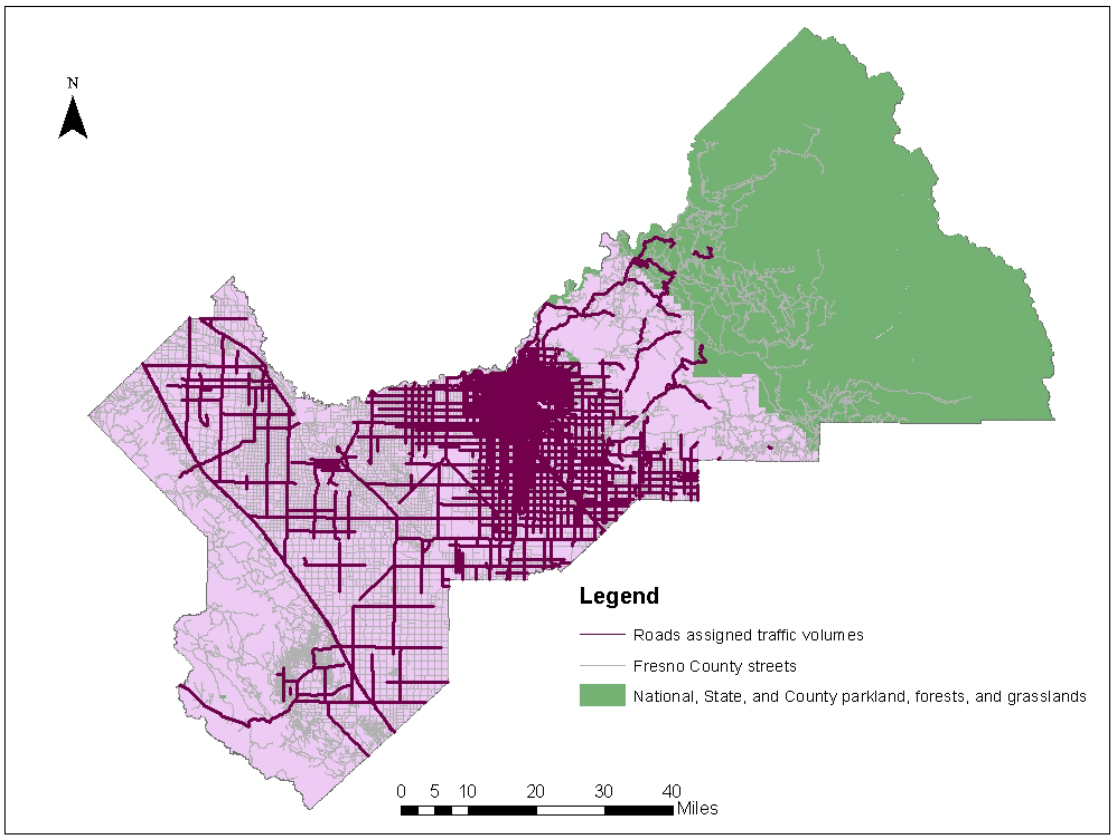
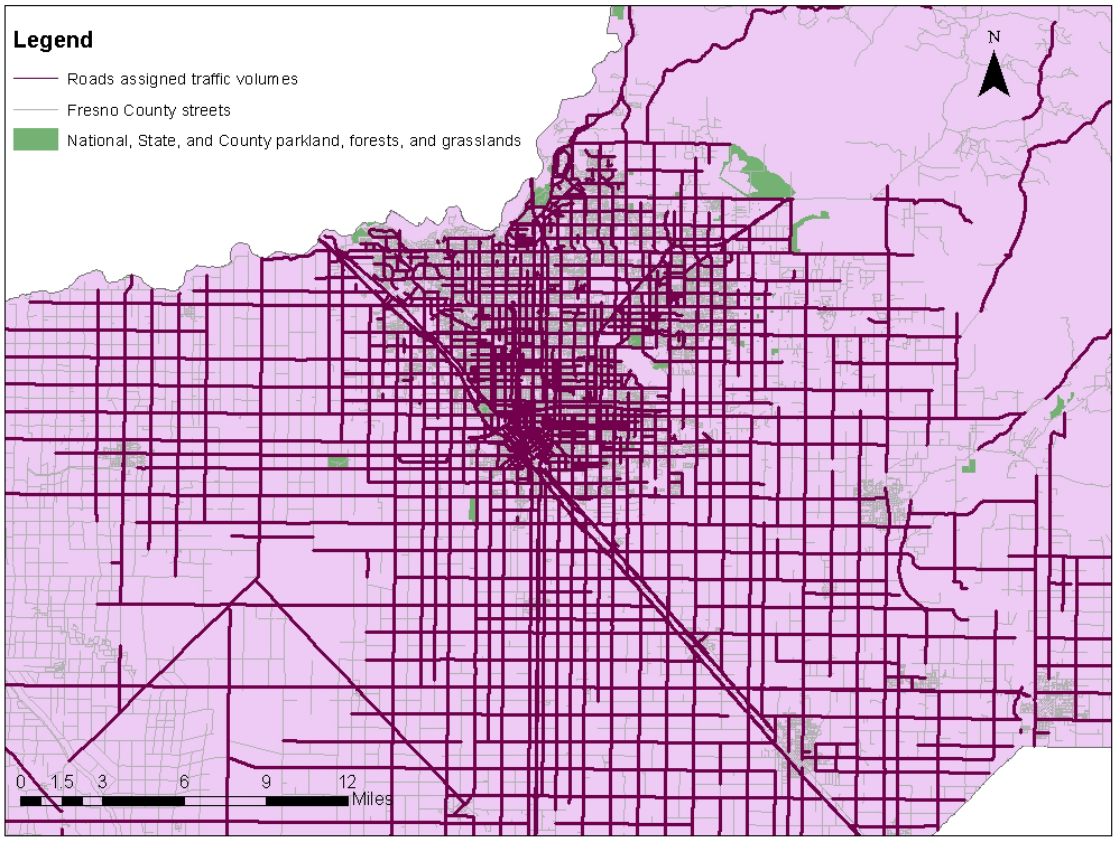
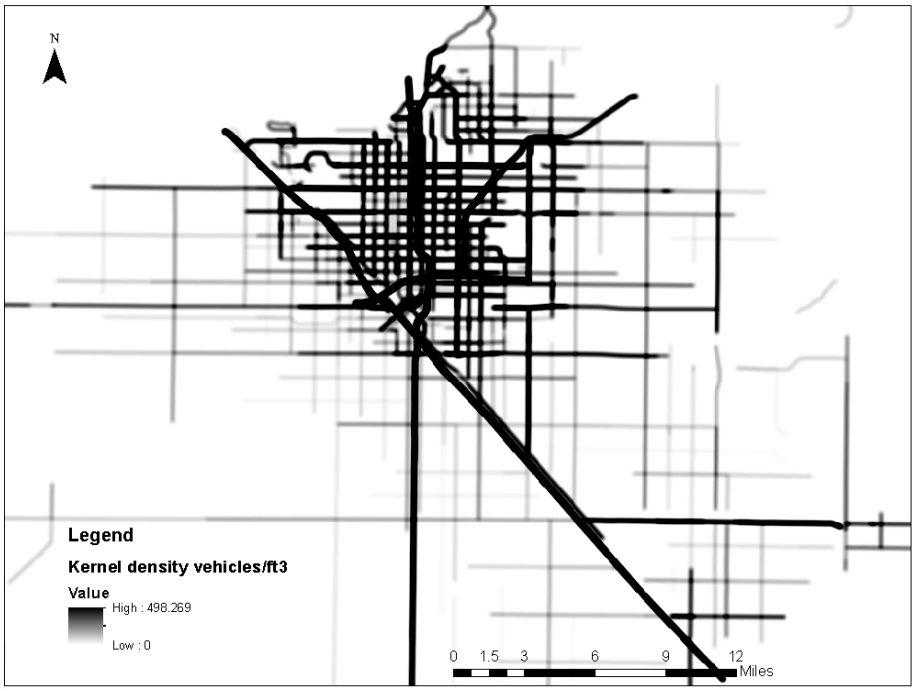


Figure 3b. Zoomed in map of Fresno County, CA FACES study area with traffic volume assignment to street network 2000-2008



Kernel densities by year illustrate heavier traffic density along the freeways and Fresno City area and lower traffic density in the rural areas (Figure 4).

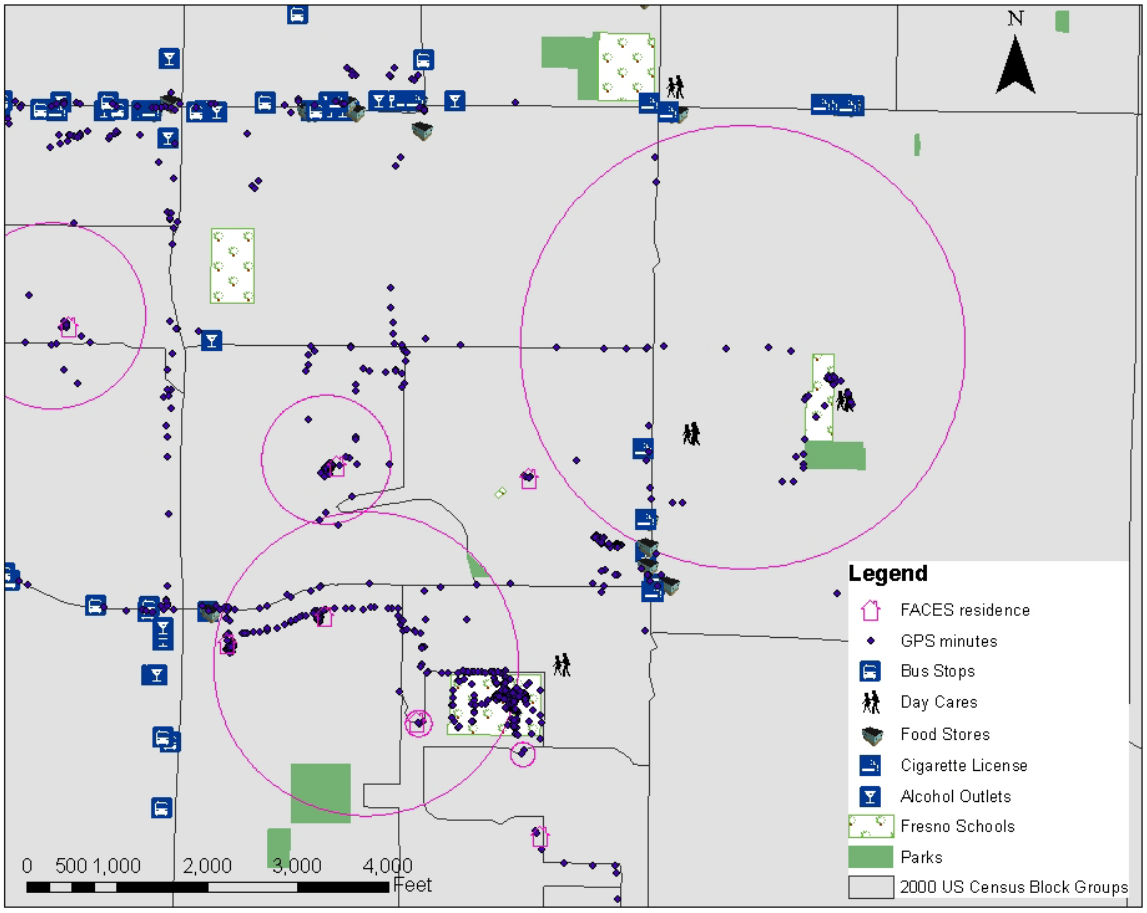
Figure 4. Example of a kernel density in the Fresno Metro Area, CA for average daily traffic in 2005



Neighborhood Assignments

Figure 5 shows the results of a sample of FACES participants' neighborhoods based on the GPS time-activity points. Each neighborhood centroid does not necessary reflect a participant's residence with the GPS defined neighborhoods. Contrary to the GPS neighborhoods, the ½ mile walking distance neighborhoods use the participant's residence point as the centroid for the circular neighborhood with a radius of ½ mile.

Figure 5. Example of GPS neighborhoods in Fresno, CA 2006-2007



Both the GPS and ½ mile walking distance neighborhoods were assigned traffic density exposures based on the year last lived at each residence location. For example, if a participant lived at 10 Main Street from 2000 to 2004, the traffic exposure will relate to the traffic data collected for 2004. Study participants moved residences throughout the nine study years; consequently, there can be more than one neighborhood for each participant. There are 518 different residences in total for the 315 study participants. There were 134 participants who completed the GPS sub-study during October 2006 to October 2007. Children wore a GPS armband and accelerometer (activity motion detector) for 5-days (with weekend span) during a 14-day panel (see Methods "Study Population").

I limited the GPS time-activity data just to those points (n=920,259) that were within 1 mile of each participant's home to minimize time spent inside a car. I wanted to capture more immediate neighborhoods and areas where a child could be on foot or on a bike. The traffic exposures are represented as an area-weighted sum and maximum value within each neighborhood. The distribution of the traffic exposures by neighborhood is displayed in Table 3. The distributions are skewed and will be used as a binary variable (low/high) in the analysis. In general, the GPS neighborhoods were smaller (median standard distance radius=524 ft) than the ½ mile (radius=2640 ft) walking distance neighborhoods, and the distributions reflect this. The walking distance neighborhoods have greater traffic exposures than the GPS neighborhoods. In the causal inference analysis, the different neighborhoods are considered with a variable marking the type of neighborhood in all modeling steps. Figures 6 and 7 show plots of the distributions. After considering where the distributions of traffic by neighborhood start to become skewed, I choose a cutpoint of low traffic as less than 15 vehicles/cubic feet for the area weighted neighborhood traffic exposures and less than 50 vehicles/cubic feet for the maximum traffic neighborhood traffic density. These low/high measurements will be used for the remainder of the analysis for the environmental justice assessment and the causal inference analysis.

Table 3. Distribution of traffic exposures (vehicles/ft^3) based on neighborhood

	Mean	5%	25%	50%	75%	95%	Min	Max	Std. Dev
Area Weighted Traffic Density for GPS neighborhoods (n=134)	9.23	0	2.00	7.83	12.46	29.25	0	51.93	8.89
Maximum Traffic Density in GPS neighborhoods (n=134)	17.67	0	5.00	14.5	24.00	50.00	0	91.00	16.16
Area Weighted Traffic Density for ½ mile walking distance neighborhoods	12.11	0.60	4.48	9.26	16.02	33.93	0	81.88	11.33

(n=518)									
Maximum Traffic Density for ½ mile walking distance neighborhoods (n=518)	40.50	4.00	19.00	29.00	51.00	106.00	0	188.00	37.51

Figure 6. Histogram of exposure distribution of the area-weighted traffic within the GPS and the ½ walking distance neighborhoods (n=652)

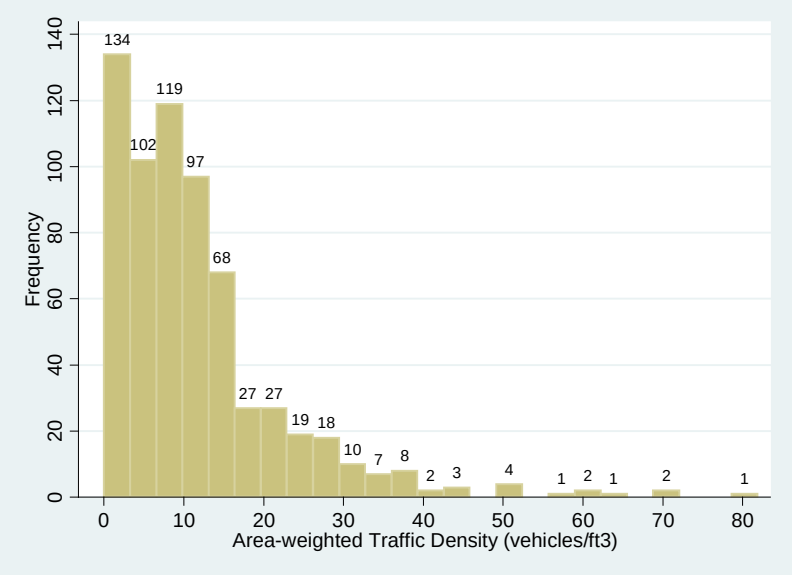
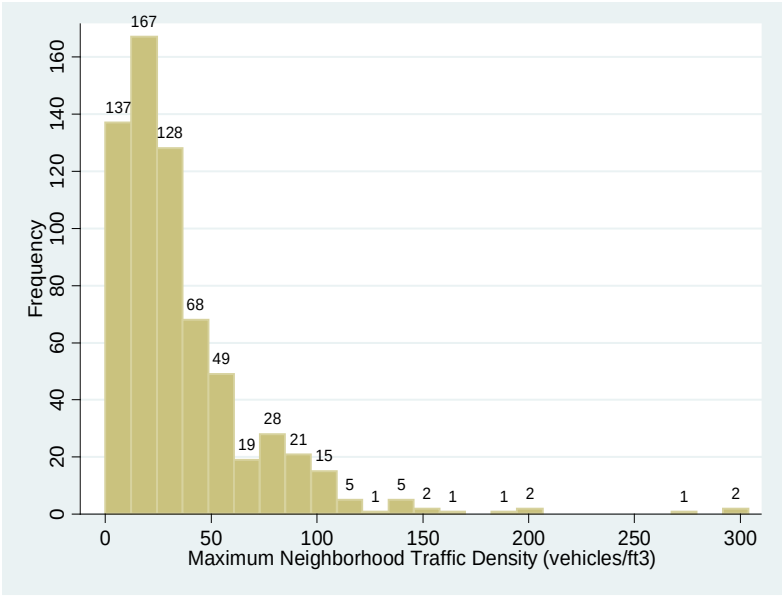


Figure 7. Histogram of the exposure distribution of the maximum density of traffic found within GPS and the ½ mile walking distance neighborhoods (n=652)

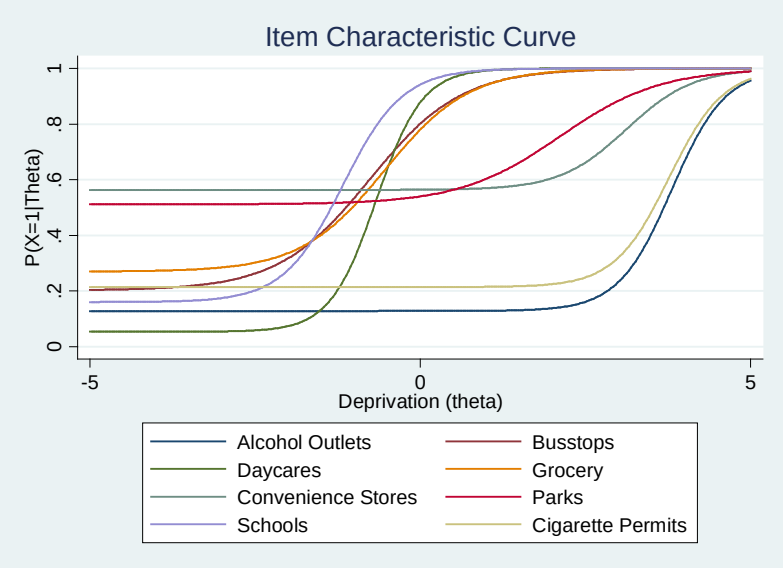


Neighborhood Deprivation

Item response scores were calculated for each neighborhood and neighborhood type. Figure 8 shows the distribution of IRT scores--with full residential history for 315 study subjects (n=518 residences), the median IRT score is 0.049 with a range from -5 to 5, SD=2.38. The median IRT score is near zero, which means that most neighborhoods are not extremely well off nor disadvantaged—most neighborhoods can be classified as in between. The item characteristic curve plotted in Figure 9 displays how each item or deprivation factor tracks with the final IRT score. Most of the neighborhoods had scores around 0, which means that these neighborhoods have a high probability of having a daycare, a grocery store, or a busstop (Figure 9). When the IRT score is equal to the estimated difficulties for each item, there is a 0.5 probability of having a positive neighborhood attribute. Cigarette permits and alcohol outlets are the "most difficult" items in this analysis. This means that the probability of having a positive neighborhood attribute is low for most participants on these items. For instance, the probability of having a positive attribute (with regard to the items, cigarette permits and alcohol outlets) is 0.5 when the IRT score is approximately 4. For items with lower "difficultly," such as daycares, the probability is 0.5 when the IRT score is slightly less than 0. Figure 10 shows that the scores correlate somewhat to US Census level poverty with a Spearman's Rho = 0.58. There are very low IRT scores (which mark high deprivation) where there are neighborhoods with low levels of poverty—we would expect to see the opposite. The IRT scores do not track with poverty either. For an IRT score of 3, the percent of poverty varies from low to high.

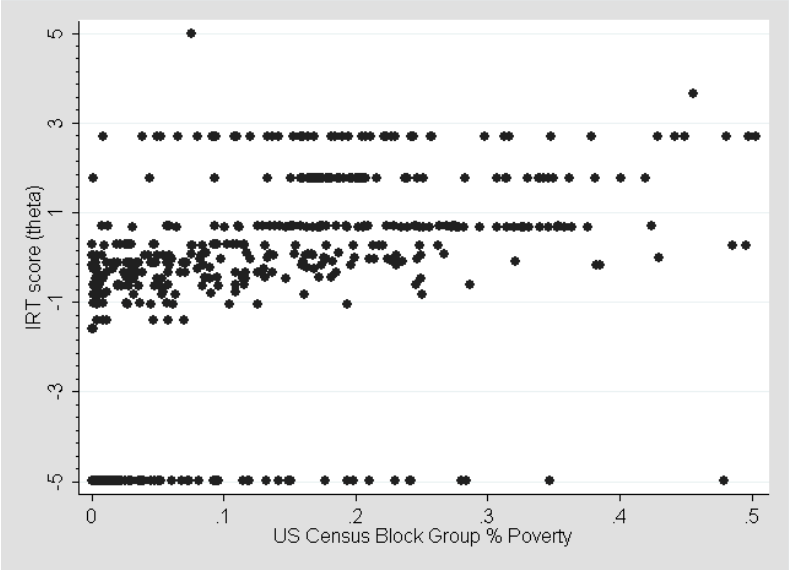
Figure 8. Boxplot of distribution of IRT scores (n=518)

Figure 9. Item characteristic curve of IRT scores and probability of participants with a positive attribute*



* When the IRT score (theta) is equal to the estimated difficulties for each item, there is a 0.5 probability of having a positive neighborhood attribute.

Figure 10. Scatterplot of Item Response Theory scores (n=518) and US Census 2000 blockgroup level percent of residents below the Federal Poverty Level



Environmental Justice Assessment

Table 4a displays the results of the index of inequity based on children in the FACES study group who have an annual family income of less than $15k and the exposure to low or high traffic. This measure of low-income (<$15k) was *a priori* selected by the FACES investigators for previous analyses. Based on the FACES study population residences, there is some

disproportanality of the low-income study participants with respect to high traffic exposures, because the index score is positive. Given that the index is on the log scale, when index values are at and near zero, there is no disproportionate clustering in space. The subgroup population is not situated near the exposure more so than other areas of Fresno, CA. The index of inequity (Table 4a) is positive for both high traffic definitions; this means that more low income study participants live near high traffic density than outside of the high density as compared to the entire FACES study population. Hence, there is some environmental injustice occurring within our study population when I use low income and high traffic exposure as my identification factors (Figure 11). The indices are small, so there is slight environmental injustice in the FACES study population.

The difference between Tables 4a and 4b relates to the characteristic used to define a subgroup for an environmental justice analysis. In Table 4a, low income is the feature that defines the subgroup. In Table 4b, I use the neighborhood deprivation score from the previous analysis to define a subgroup. With high neighborhood deprivation as the defining subgroup characteristic instead of low income, there is no inequity as measured by the index (Table 4b). On the contrary, the negative scores reveal that more FACES participants who are in high deprivation neighborhoods live farther away from high traffic areas. This could be due to the fact that many of the variables that describe deprivation in this analysis are situated in areas where we have more cars (e.g. convenient marts and liquor stores are often in busier areas, while parks may not be as frequent in such areas). The interpretation of these results is furthered in the discussion.

Table 4a. Environmental justice index of inequity† results for FACES participants with income and residential data (n=302) at baseline

Traffic Exposure	Low Income FACES Participants, n=62(%)	95% Confidence Interval‡	Index of inequity
Exposed to High Traffic ($\geq$ 15 vehicles/ft^3 with area weighted method)	26 (41.9%)	(18.38, 33.62)	0.31
Exposed to High Traffic ($\geq$ 50 vehicles/ft^3 with maximum neighborhood method)	18 (29%)	(10.99, 25.01)	0.15

† $F_i = \log(Z_i/T_i)$ where F is the inequity experienced by low income FACES participants *i* and Z is the fraction of the total low income FACES population near high traffic pollution and T is the fraction of the total subgroup FACES population within Fresno

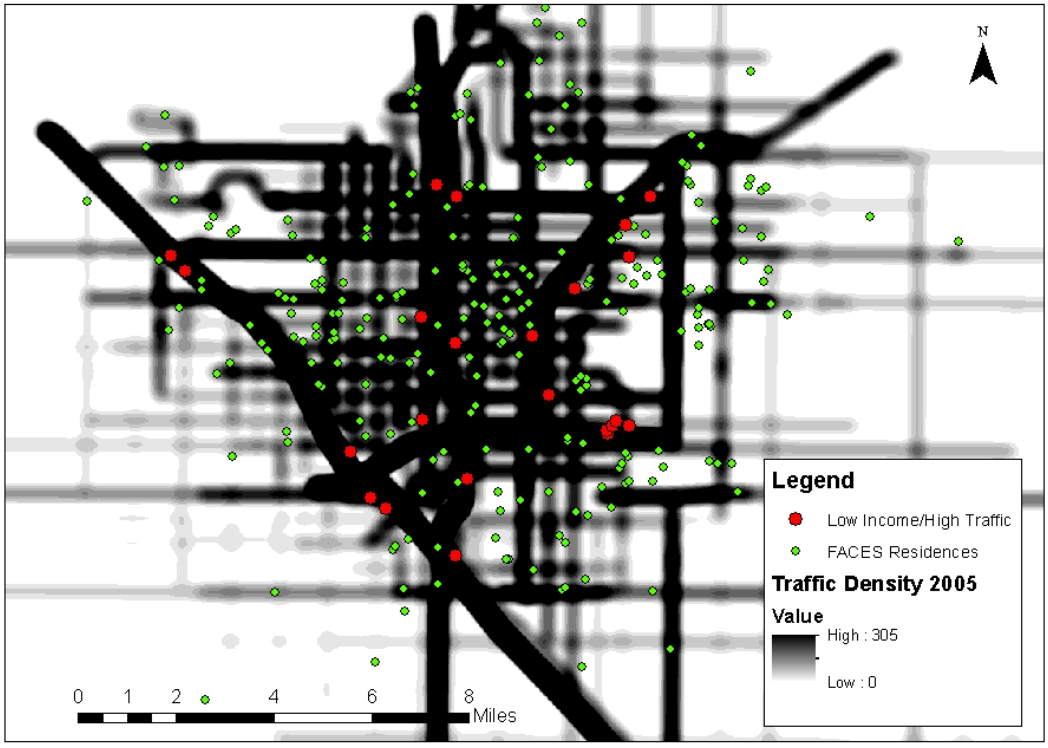
‡Based on a binomial distribution

Table 4b. Environmental justice index of inequity† results for FACES participants based on deprivation scores, where high neighborhood deprivation is defined by an IRT score $\leq$ -1

Traffic Exposure	FACES Participants living in high deprivation neighborhoods, n=103	95% Confidence Interval‡	Index of inequity
Exposed to High Traffic ($\geq$ 15 vehicles/ft^3 with area weighted method)	15 (14.6%)	(7.98, 22.02)	-0.119
Exposed to High Traffic ($\geq$ 50 vehicles/ft^3 with maximum neighborhood method)	9 (8.7%)	(3.38, 14.62)	-0.341

† $F_i = \log(Z_i/T_i)$ where F is the inequity experienced by FACES participants *i* living in high deprivation neighborhoods and Z is the fraction of the total FACES population in high deprivation neighborhoods near high traffic pollution and T is the fraction of the total subgroup of FACES population within Fresno
‡Based on a binomial distribution

Figure 11 Map of Fresno County, CA with 2005 kernel traffic density and study participants exposed to high traffic densities with less than a $15k household income



Causal Inference Analysis

The FACES study population consists of children who were on average 8 ½ years old when they began the study (Table 5). One fifth of the study participants came from low income families. About 40% were Hispanic. Half of our participants had mild persistent asthma, and one quarter had moderate to severe asthma.

Table 5. Descriptive statistics for FACES participants

Characteristic	Full cohort (n=315)
Mean Age at baseline [S.D.]	8.5[1.7]
Male (%)	56.5
Income Less than $15,000 (%)	20.4
Home Ownership (%)	56.5
Health Insurance (%)	95.9
Hispanic (%)	39.7
Non-hispanic white (%)	41.9
African American (%)	16
Mean # of Panel Visits Completed [s.d.]	8.4 [5.0]
Skin-test positive to at least one antigen (%)[a]	62.7
Mild intermittent asthma (%)[b]	28.2
Mild persistent asthma (%)[b]	47.6
Moderate or severe asthma (%)[b]	24.1
Use inhaled steroids (%)	73.0
Oral prednisone, last 12 months (%)	37.5
%FEV_1 < 80 % predicted[d]	17.5
%FEF_{25-75} < 70 % predicted[d]	26.2

For the causal inference analysis, I considered variables that were related to both traffic exposure and lung function as confounders. I also considered variables that were known to be strong associated with the outcome (lung function), but maybe had proxy pathways or worked through an unmeasured variable to be related to traffic exposure. The variables in Table 6 were selected *a priori* to have a relation with lung function based on the literature and previous analyses with the FACES cohort. The bivariate associations with these variables and FEV_1 are displayed in the Table 6 by order of significance.

Table 6. Candidate variables considered for analysis shown with Chi-squared p-values for bivariate associations by ID with the outcome for annual mean FEV_1 of panel visits

Variable	Description	Chi-squared Z-value	P-value
height_c	Annual mean height cubed (cm)	16.15	0.000
mheight	Annual mean height (cm)	15.50	0.000
mage	Mean age	14.35	0.000
Race4	0=other race, 1=Asian	-13.69	0.000
mweight	Mean weight (lbs)	9.14	0.000
Year08	Panel test in 2008	8.57	0.000
Year04	Panel test in 2004	-6.66	0.000
junip_p	skin test positive to juniper	-5.49	0.000
Year01	Panel test in 2001	-5.39	0.000
Year03	Panel test in 2003	-4.93	0.000
Year02	Panel test in 2002	-4.26	0.000
ncoh8	Child entered study after 7/1/2004	-4.18	0.000
Gps_hood	0=traffic exposure data from ½ mile walking distance neighborhood, 1=traffic exposure data from GPS based neighborhood	3.87	0.000
race2	0=other race, 1=black	-3.67	0.000
cedar_p	skin test positive to cedar	-3.39	0.001
ncoh6	Child entered the study between 7/1/2003 and 12/31/2003	-3.24	0.001
Year05	Panel test in 2005	-2.92	0.003
asthle2	Diagnosed with asthma before the age of 2	-2.64	0.008
Year07	Panel test in 2007	2.63	0.008
mosmk_pr	1=mother smoked during pregnancy	-2.46	0.014
atopy	skin test positive to at least one allergen	2.35	0.019
ncoh1	Child entered study before 6/30/2001	2.31	0.021
ncoh3	0=other 1=baseline visit was between 1/1/2002 and 6/30/2002	2.30	0.022
atopy2	Skin test positive to at least one allergen or had a	2.25	0.024

	severe reaction		
ncoh7	child entered study between 1/1/2004 and 6/30/2004	-2.22	0.026
olive_p	Skin test positive to olive	1.95	0.051
income1	1=inclome below $15K, 0=otherwise	-1.94	0.053
income4	income > $50K	1.93	0.054
gina3	0=other GINA score, 1=Moderate/Severe severity	-1.82	0.068
everhosp	was ever hospitalized for asthma at baseline	-1.79	0.073
cat_p	skin test positive to cat dander	1.73	0.084
inccat	0= income> $30k, 1=Income <$30k	-1.53	0.127
race1	0=other, 1=hispanic	1.47	0.142
evericu	0=no, 1= ever spent time in ICU for asthma assessed at baseline	-1.46	0.145
male	0=girl, 1=boy	1.38	0.169
alt_p	0=negative, 1=positive for alternaria allergy	1.30	0.195
grass_p	skin test positive to grass	1.28	0.202
rye_p	Skin test positive to rye	1.24	0.216
brstfed	0=not breastfed, 1=breastfed	1.13	0.261
penc_p	skin test positive to penicillin	-0.99	0.322
clado_p	skin test positive to cladosporium	0.85	0.395
gina2	0=other GINA classification 1=GINA classified mild persistent asthma	0.84	0.398
mite_p	skin test positive to mites	0.81	0.419
numhosp	number of hospitalizations at baseline	-0.68	0.496
gina1	0=other GINA score, 1=Mild intermittent severity (lowest severity level)	0.68	0.499
asth_fa	0=father does not have asthma 1=father has asthma (baseline)	-0.56	0.576
race3	0=other 1=white	0.53	0.596
privet_p	skin test positive to privet	-0.45	0.650

ncoh5	0=other 1=baseline visit was between 1/1/2003 and 6/30/2003	-0.36	0.721
mug_p	skin test positive to mug worm	0.34	0.736
income2	0=income<$15k or >$30k, 1=income between $15 and $30K	0.33	0.738
lbw	0=not low birthweight, 1=lowbirthweight	0.22	0.822
lbwprm	low birth weight or born prematurely	-0.22	0.828
dog_p	skin test positive to dog dander	0.18	0.855
income3	1=income between $31 and $50K, 0=otherwise	-0.18	0.858
asth_mo	0=mother does not have asthma 1=mother has asthma (baseline)	0.17	0.868
prem	0=not premature, 1=born prematurely (>3 weeks early)	-0.13	0.893
ncoh2	child entered study between 7/1/2001 and 12/31/2001	-0.13	0.896
ncoh4	child entered the study between 7/1 and 12/31/2002	-0.10	0.917
Year06	Panel test in 2006	0.07	0.942
oak_p	0=negative, 1=positive for oak allergy	0.02	0.987

To understand if neighborhood deprivation modifies the relation between traffic exposure and lung function, I ran three different analyses. Analysis 1 does not take neighborhood deprivation into account in the modeling. Analyses 2 and 3 are based on subsets of the data. Analysis 2 is only performed on those study participants who lived in areas of high neighborhood deprivation (IRT score $\leq$ -1). Analysis 3 is only performed on those study participants who do not live in high deprivation or an IRT score > -1. The candidate variables (Table 6) were entered into the DSA for model selection for the Q model, the g or treatment model and the censoring or missingness model for all analyses. The DSA arguments were set to a maximum model size of 12 variables, a maximum sum of power of two, and a maximum set on interactions of two variables. The exposure, high traffic as defined by 15 or more vehicles per cubic foot, was forced into the Q model. The DSA results are listed in Table 7. Each model was run 10 times with 10 different user seeds. The most frequently selected models (displayed in Table 7) were used in the final analyses, except for Analysis 3 where the most frequently selected model resulted in small cells. For Analysis 3, the model selected 6 out of 10 times had only 3 study participants in one of the combinations, so I used the next best model, which was selected by the DSA 2 out of 10 times and is displayed in Table 7. For all of the analyses, the DSA did not select a model for the data missing an outcome, which means that there were no variables in the data that predicted missingness. All Q models contained height, but the Q models in Analysis 1 and 3 contained age x height in the final models to best predict FEV_1. The

treatment or g models all had different combinations of variables to predict the treatment or high traffic exposure.

Table 7. Models selected for analysis by the DSA

Analysis	Q model	g model	Censoring model
1) Exposure = neighborhood traffic area-weighted average Outcome = average panel lung function (FEV_1) (n=538)	(High traffic) + (mean age x height^3)	(Father asthma diagnosis) + (skin test positive for cladosporium) + (mother smoked during pregnancy) + (income >50k)	No model selected
2) Exposure = neighborhood traffic area-weighted average Outcome = average panel lung function (FEV_1) *Modifier* = High Neighborhood Deprivation (IRT score $\leq$ -1) (n=103)	(High traffic) + (height^3)	(Father asthma diagnosis)	No model selected
3) Exposure = neighborhood traffic area-weighted average Outcome = average panel lung function (FEV_1) *Modifier* = Low Neighborhood Deprivation (n=435)	(High traffic) + (mean age x height^3)	(skin test positive for cladosporium)+ (mother smoked during pregnancy)+(Hispanic race) + (Father asthma diagnosis)	No model selected

I plotted each g model to test for ETA violations. In general, the g models all had low predicted probabilities for high traffic exposure. The g model for Analysis 1 had the most

observations and also the most variables in the DSA selected model, so the plot (Figure 12) shows the greatest variability of values as compared to Analyses 2 and 3. Figure 13 shows that Analysis 2 has fewer observations and also fewer terms selected by the DSA for the g model. The predicted probabilities are low, but the values do not always predict 0, or have no probability of having a high traffic exposure. Analysis 3 has a similar ETA plot to Analysis 1 (Figure 14)—likely because Analysis 3 is only missing 103 study participants from stratification and the g-models for both analyses are similar.

Figure 12. Plot of the Experimental Treatment Assignment with the predicted probability of high traffic exposure and log odds of high traffic exposure for analysis 1 (labels=n)

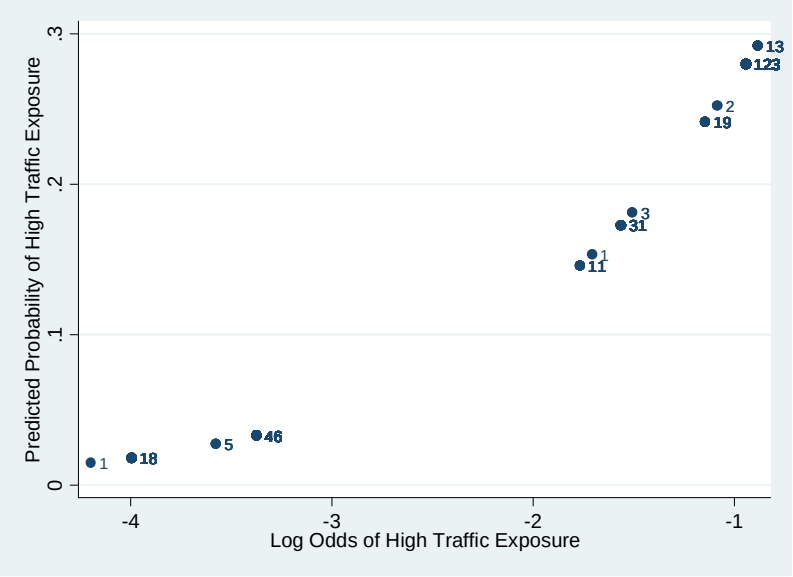


Figure 13. Plot of the Experimental Treatment Assignment with the predicted probability of high traffic exposure and log odds of high traffic exposure for analysis 2 (labels=n)

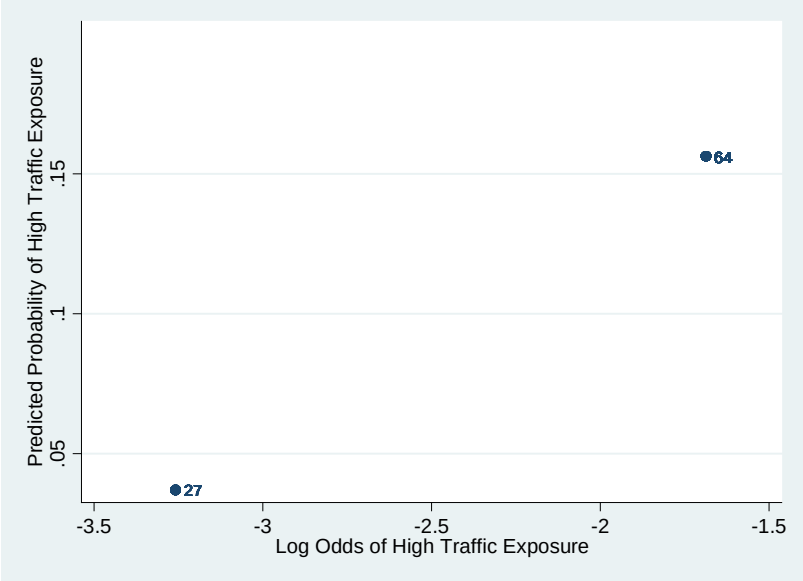
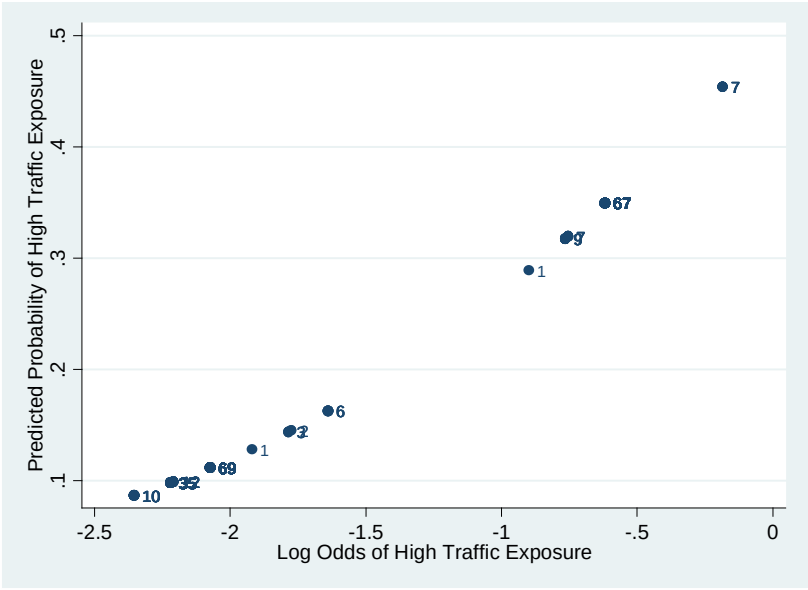


Figure 14. Plot of the Experimental Treatment Assignment with the predicted probability of high traffic exposure and log odds of high traffic exposure for analysis 3 (labels=n)



The final TMLE estimates are presented in Table 8. The risk difference for Analysis 1 without stratification for neighborhood deprivation is -0.233 (95% CI -0.338, -0.129), which means that the mean lung function change from low traffic to high traffic is an average decrease of 0.233 L in FEV_1. This says that children who are exposed to lower neighborhood traffic have greater lung function than children who are exposed to higher neighborhood traffic. When I compare the TMLE results by neighborhood deprivation (high deprivation is an IRT score $\leq$ -1), there is not enough data to estimate the mean population lung function based on exposure and stratified by neighborhood deprivation. The estimate is imprecise, but the trend when compared

to Analysis 1 and 3 is that the reduction in lung function is lessened for the FACES group that lives in highly deprived neighborhoods. The rest of the FACES group (Analysis 3) has a result similar to the full group. The risk difference is -0.198 (95% CI: -0.344, -0.053), which means that the mean lung function change from low to high traffic among those participants not living in highly deprived neighborhoods is an average decrease of 0.198 L. Traffic appears to be separate from neighborhood deprivation and a more important factor to consider. The exposure profiles of the stratified groups are quite similar (Figure 15a and b). Furthermore, with regard to the IRT results, high traffic exposure is not that different by income (Figure 16a and b).
The expected percent reduction in lung function (the risk difference divided by the average expected lung function when high traffic exposure is zero) for Analysis 1 is 12%. Therefore, the marginal decrease in lung function is 12% from exposure to high traffic among FACES study participants.

Table 8. Risk differences for each analysis

Analysis	Risk Difference from High traffic ($\geq$ 15 vehicles/ft^3) to Low traffic (< 15 vehicles/ft^3)	95% Confidence Interval	[Estimated Variance] P-value
1) Exposure = neighborhood traffic area-weighted average Outcome = average panel lung function (FEV_1) (n=538)	-0.23323	(-0.338, -0.129)	[0.003] p<0.0001
2) Exposure = neighborhood traffic area-weighted average Outcome = average panel lung function (FEV_1) Modifier = High Neighborhood Deprivation (IRT score $\leq$ -1) (n=103)	-0.11355	(-0.433, 0.206)	[0.027] 0.486
3) Exposure = neighborhood traffic area-weighted average Outcome = average panel lung function (FEV_1) Modifier = Low Neighborhood Deprivation (IRT score > -1) (n=435)	-0.19826	(-0.344, -0.053)	[0.006] 0.008

Figure 15a. Distribution of traffic exposure among FACES participants who do not reside in highly deprived neighborhoods

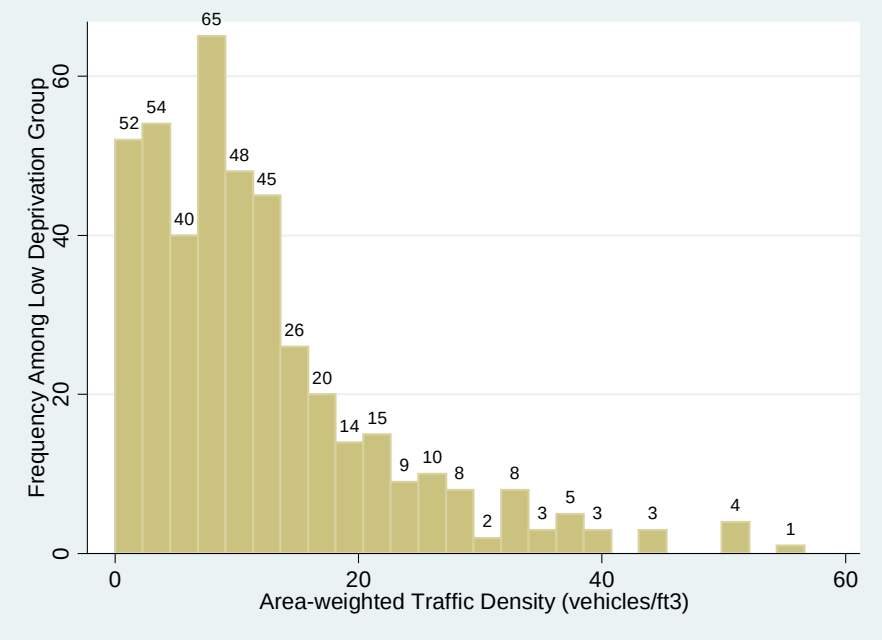


Figure 15b. Distribution of traffic exposure among FACES participants who reside in highly deprived neighborhoods

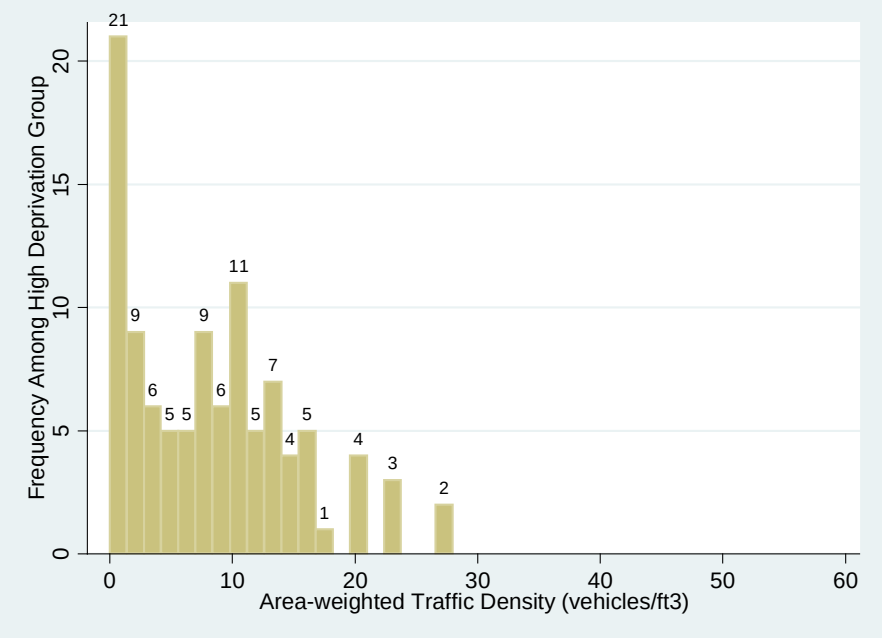


Figure 16a. Distribution of traffic exposure among FACES participants who are low income (family annual income of <$15k)

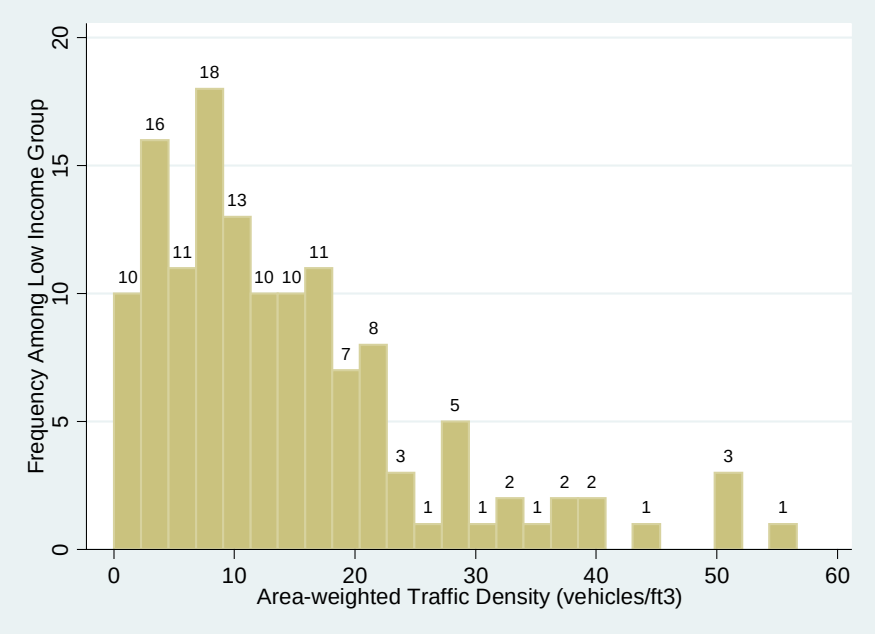
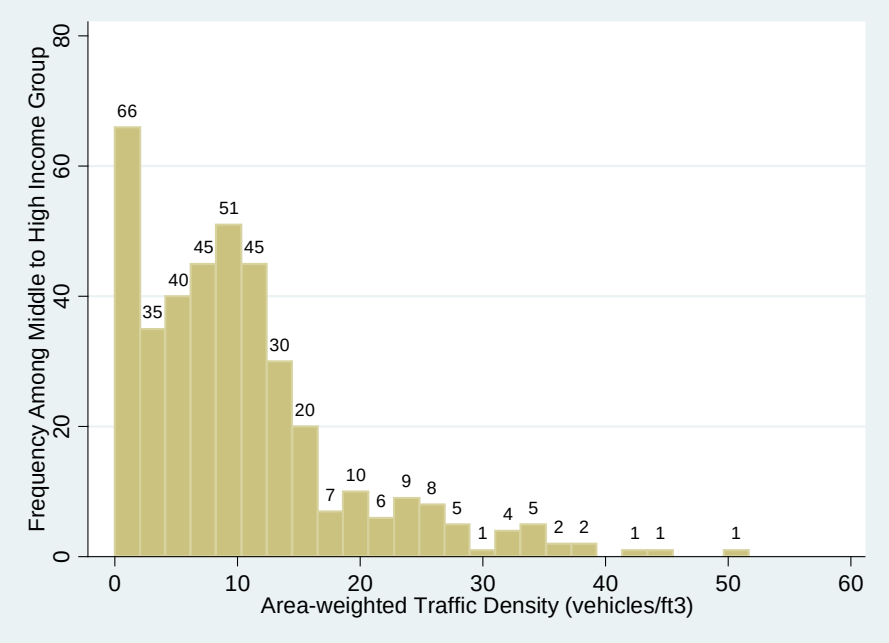


Figure 16b. Distribution of traffic exposure among FACES participants who are not low income (family annual income of $15k or greater)



Chapter 4: Conclusion

Traffic Exposure Assessment

The traffic data that I collected for this analysis are the most comprehensive that this study area of Fresno, CA has to date. Previous health-related work in the FACES project have used sparse traffic data from one source (e.g. CalTrans freeway data where n~200 data points per year for Fresno County) to approximate: 1) traffic exposure for the FACES study population;[106] and 2) traffic density for a land use regression model to estimate polycyclic aromatic hydrocarbon exposure.[127] The CalTrans data account for traffic on freeways, highways, arterials and some major collector roads; thus, most local roadways are not part of the data. While heavier traveled highways are the most important contributors to ambient air pollution, the secondary and local roads may be more important to classify immediate exposures near residences and to gain more complete coverage for larger study areas. Recent studies that model traffic for sprawled metropolitan areas, such as Los Angeles, CA, have traffic data from local roadways (n~8000) contribute almost half of the data points to a traffic model for land use regression analyses.[118,128] Other studies are vague with regard to or do not disclose how many data points were used to estimate traffic density exposures over large study areas.[80,113] Moreover, these studies sometimes use traffic data that were collected during years different from their health study.[113]

I aimed to get the best available data for Fresno County for all years of our study. The data span different seasons, years, roadway types, and agencies; they also provide an annual snapshot of traffic for local roadways. However, the data are not without their limitations. Since both traffic exposure and lung function are annual averages, there could be some exposure misclassification. For instance, the traffic exposure estimate might not always represent the exposure that preceded the outcome. While this is a problem and ought to be considered in the interpretation of the results, this misclassification would be worse if I used annual traffic data to look at short term effects, such as acute, daily changes in lung function from annual averages of traffic. The analysis was meant to match the data—a comparison of annual averages. There are likely traffic measurements that were collected after lung function spirometry tests, but both measurements are the result of means. The data are the best estimate of a mean exposure that a person has related to pulmonary function measurements. Since I do use averages for both measurements, the misclassification of temporality is lessened with the inclusion of multiple data points. For the county and city agency data, I use 24 hour counts within a study year to approximate an annual average. A one-day count is most likely not indicative of the traffic experienced at that location for the whole year. However, the one-day count is an unbiased but highly imprecise estimate. Also, the county and city agency data do not have traffic information on all roadways in their jurisdiction--the data are incomplete. Of the data that I collected, I had excellent geocode matches. Less than 5% of the data were unable to be matched (Table 2). Finally, most of our study subjects lived within the city of Fresno, where we had the best road network coverage (Results Figure 1b and 2b).

Neighborhood, Item Response Theory Model and Environmental Justice

Single aggregate variables are inadequate to describe the degree to which people live in advantaged or disadvantaged neighborhood.[7] Using the area-weighted neighborhood method to classify neighborhood characteristics and with Item Response Theory, I allow for consideration of this complexity. I can look at a number of factors and summarize them quantitatively to capture deprivation/advantage with varying definitions of neighborhood in terms of quality and

characteristics. This classification is based on a pseudo-individual neighborhood, not grouped data. The individual neighborhoods partially solve the problem of non-identifiability or small cells without overlapping covariate structure among exposure groups. While the deprivation measures and IRT scores do not track completely with census level poverty, there is no gold standard measurement for neighborhood deprivation. Each agency or research group ought to use the measures most appropriate for their analysis though. In the case of health analyses, we have found that deprivation is something complex that can affect health outcomes via individual and aggregate mechanisms. Thus, sometimes the easiest data to obtain that describe a neighborhood, such as US Census data, might not be the most appropriate to use in a health analysis. In my analysis, I was interested in both the contextual and integral factors in neighborhoods of our study participants in Fresno, CA. The US Census only provides aggregated contextual factors, which would portray part of the picture of neighborhood deprivation. The additional factors that I collected along with the IRT analysis helped provide a composite factor and wider view of neighborhood deprivation. This is more powerful and inclusive than say, the use of US Census measured poverty to describe neighborhood deprivation. The field of public health benefits from a wider lens when we do not have a go-to measure for a latent variable such as deprivation.

With environmental justice assessments of the impact of traffic on health, we often must choose a characteristic that identifies collectively a social group or a sensitive population though. However, the dilemma is to identify which characteristics best describe populations in need of environmental protection? For my environmental justice analysis, I chose neighborhood deprivation and low income. Deprivation describes a population living in an environment that is in need of positive amenities and possibly has more negative influences (e.g. liquor stores) and, perhaps lacks, the power to change their exposure to traffic. Individual-level low income describes a population that may have financial difficulty needed to make a change to their exposure to traffic (e.g. property can be cheaper near busy freeways and low income families may not be able to afford to move). With deprivation as the criterion to define a sub-group, there was no environmental injustice occurring; I see the opposite. The index is negative, which shows that study participants who live in areas with high neighborhood deprivation have lower traffic exposures. This could be due to the location of amenities; there are more positive neighborhood influences (e.g. grocery stores, daycares, busstops, etc.) in areas that have denser traffic. This analysis could be problematic, because area-level attributes were used to define the subgroup and also define the exposure. We know that nearby spatial attributes can be autocorrelated, so it is possible that the index is depicting area level correlations. With low income as the definition, I use an individual characteristic to make a comparison to an area-level attribute. Individual income is a more common factor to define a disenfranchised sub-group, and with this definition there is slight environmental injustice in the study population. The index of inequity shows a positive score, which means that study participants in the low-income group are disproportionately located near high traffic densities.

The results of the environmental justice analysis tell two parts to the same story. First, living in what I define as a deprived area in Fresno, CA likely means being away from services and the associated higher traffic. Second, deprivation does not track well with poverty, and the environmental justice results with income reveal that poorer families reside in areas with heavier traffic. Thus, in Fresno, CA, wealthier families in our study population live in more rural areas with lower traffic density and fewer amenities and services.

There are several limitations to this work. The IRT for deprivation might not be a good model for what we consider neighborhood deprivation in rural areas and items that reflect contrary deprivation definitions might be found in the same place (e.g. grocery stores are positive neighborhood influences, but off-sale alcohol outlets might be found at the same or a nearby location). Also, with any modeling, the IRT could be missing unmeasured data that better reflect deprivation. Spatial positional accuracy could affect the IRT scores as well. If the spatial inputs, such as parks and daycare locations, are grossly wrong, then the IRT score will be biased. However, since the spatial variables were transformed to a binary indicator (i.e. geographic attribute present or not present), the error will likely only cause bias around the smaller values to the extent that values cluster around the cut point.

The GPS data I use to define a neighborhood is only a snapshot of how a study participant spends his or her time, so, as with the walking distance defined neighborhoods, the data might not be representative of a study participant's actual neighborhood. However, this measurement might not be yield a precise estimate of where a study participant roams near his or her home, but it is an unbiased estimate of where a study participant goes nearby. My approach is the first neighborhood study in the literature that uses GPS to define neighborhood boundaries for individual study subjects. Most of epidemiology relies on aggregated data with boundaries from the US Census. The census does not track people to see objectively where they go near their residence. Rather, the census neighborhoods are based on population size and physical boundaries.

The GPS data are not without their limitations. They do contain error; the accuracy of the device is that 95% of the time the indicated location will be within 6 m or 20 ft. Most of the neighborhoods were much larger than 20 ft—the median GPS neighborhood had a radius of 524 ft. Again, the error in the device will likely bias the results at very small values and very few observations. Since I use minute-by-minute GPS points over a period of 5 days, the error is likely minimized.

I am limited in this analysis in that I cannot make any individual level inferences about traffic, deprivation and lung function, because this model does not look at how deprivation and traffic directly affects children in our study. To calculate deprivation and traffic more precisely, individual questionnaires could be administered along with backpacks that monitor air quality around the individual--our model cannot capture a study participant's perceptions of their neighborhood deprivation, individual traffic-related air pollution and exposures inside of the home. Other limitations to this work are that we assume the US Census 2000 measures to be constant over our study time period 2000-2008, children lived at their reported residence, and the other GIS deprivation variables are constant from 2000-2008. Since the US endured an economic recession from December 2007-June 2009, the geographic variables must be interpreted with caution. The housing downturn and unemployment increases likely changed US Census characteristics and many of the other geographic factors collected for this analysis. However, the FACES group ended data collection in September 2008; therefore, the US recession coincides with less than a year of the study. The US Census describes the difference between many years of data and just one year as a trade-off of accuracy and currency.[129] The multiyear estimates will have smaller errors, but they can miss relevant demographic and geographic changes that occur over short time periods or before the latest census data are available.

Data and relationships between data can be influenced by the size and/or shape of the units from which the data are reported. [94] This issue, known as the "modifiable areal unit problem" (MAUP) [95,96] is critical in the use of spatial analysis to inform community level policies and interventions. The identification of an adequate neighborhood area should be considered vis-à-vis the disciplinary framework in which the issue is approached. For example, the American Institute of Architects Architectural Graphic Standards, as well as city and regional planners continue to use the neighborhood definition developed by Perry in the 1920s [97] as a planning unit with a five-minute walking radius.[92] The MAUP has no definitive solution but remains a challenge to spatial research that should not be overlooked. Researchers need to consider both the appropriate scale and aggregation options when using spatial boundaries. My previous work[98] illustrates the MAUP—as I changed neighborhoods from the census defined areas to the neighborhoods with walking distances, I saw significant differences in exposures (crime rate, alcohol outlet rate, etc.) and in general, the differences wane as the buffer sizes decrease. This is expected and should highlight the need for researchers to carefully consider neighborhood definition and ask: what is the most appropriate size and shape of neighborhoods for this research?

Although much of the literature on neighborhoods and health has shown associations between the built environment, social factors, and health, scant attention has focused on neighborhood definition. My method could be used in other locations to understand how features of the built environment are associated with health outcomes or the differences in harmful exposures between those who reside in neighborhoods where deprivation is high and those who do not.

Causal Inference

My hypothesis that children with asthma who are exposed to high levels of traffic experience worse lung function than children with asthma who are exposed to lower levels of traffic was correct. The findings indicate that neighborhood exposure to traffic adversely affects lung function among the FACES cohort of children with asthma. The magnitude of the estimated effects changed when the data were stratified by neighborhood deprivation as defined by the IRT scores. However, the results were no longer significant for the high deprivation group (Analysis 2). The estimated effects for the stratified analyses both show reductions in lung function, although the estimates are less than the full group in Analysis 1. The estimates are derived from different models (Table 7) and cannot be interpreted as part of the whole group in Analysis 1. Interpretation of the reduction in lung function for the stratified analyses is that there is not enough data to understand if neighborhood deprivation modifies the effect of traffic on lung function. There may be a true difference between strata of deprivation, but I would need more power to be able to distinguish a significant difference. The results for the unstratified analysis do show an effect estimate of reduced lung function (12%) among children exposed to high levels of neighborhood traffic. A 12% reduction in lung function is a considerable adverse affect for a child, especially a child with asthma. The reduction is relevant to other studies of chronic exposures and the effect of chronic exposure to a pollutant on lung function. The potential importance of this reduction can be seen via a comparision with long-term exposure to ozone. Lifetime exposure to ambient ozone is shown to reduce lung function among adults as much as 38% and 37% with a comparison of FEF25-75/FVC in the highest and lowest quartiles among men and women respectively.[130] (Lung function decreases in children with asthma who

are exposed to secondhand smoke have been documented at -4.7% (95% CI: -9.9, 0.5)) for FEV_1, -8.5% (95% CI: -15.2, -1.9), and -3.0% (95% CI: -4.9, -1.1) for FEV_1/FVC.[131]
In the model selection procedure, the type of neighborhood, GPS based or walking-distance, was not selected by the DSA as a confounder for the effect of traffic on lung function. Because I did not measure how children are individually affected by traffic, these findings only pertain to the population of children with asthma who live in neighborhoods with high levels of traffic. The traffic could be a marker or proxy for something else occurring in the neighborhood or it could be a proxy for higher ambient air pollution due to traffic.

Few confounders affected the outcome and exposure in the DSA model selection for the Q, g and censoring models. Most notably, age and height were strong predictors of lung function in this analysis and myriad other studies have documented a similar relation.[132-134] The effect estimates in this study can be interpreted as population-level effects rather than effects conditional on many confounders. Conditioning on "nuisance" confounders limits the interpretation of traditional models, which can only provide an effect estimate whose validity is dependent on adjustment for confounders. One strength of this analysis is the use of the machine-learning DSA algorithm to determine which covariates, among those associated with lung function, should be retained in the model. This algorithm requires no *a priori* model form and uses cross-validation to identify the optimal model across the entire model space of a linear polynomial whose parameters can be set by the user. Moreover, I report only models that gave the ideal request in at least four of 10 independent runs of the data. This approach reduces the probability of over-fitting the model.

Several caveats need to be considered in the interpretation of these data. We did not estimate individual-level exposures for traffic or traffic-related pollutants. Thus, we cannot estimate the degree to which the effects I report here are independent of actual exposure to traffic or other pollutants. Though previous work with the FACES cohort and traffic related air pollution indicate strong correlations with polycyclic aromatic hydrocarbons (PAHs), $PM_{2.5}$, CO, NO_2 and elemental carbon, which reflects a strong contribution of the pollutants from traffic.[135] My finding of reduced lung function among children with asthma who are exposed to high levels of neighborhood traffic is similar to other studies on health and traffic. Recently, researchers in Windsor, Ontario found a trend among children with asthma for decreasing lung function associated with exposure to traffic density within 200m from a residence.[136] This study used refined traffic counts to approximate a surface of exposures. Short-term exposure to primary traffic pollutants in ambient air (as measured by a network of ambient monitors) in Atlanta, GA were shown to increase emergency department visits for asthma.[137] A log-unit change in residential traffic counts (based on a detailed group of sites that collected data over 20 years and validated by video monitoring) in Perth, Austraila was responsible for a 24% increase in emergency department contacts among adults with asthma.[138] A sister study to FACES in the San Joaquin Valley of California shows that prenatal exposure to higher traffic density (based on CalTrans freeway and major roadway traffic counts) results in an increased risk for term low birth weight also with a TMLE analysis.[139]

In short, this dissertation aimed to provide a more comprehensive assessment of neighborhood characteristics and exposure to traffic and to investigate the relation that those factors had with lung function. Neighborhood deprivation was not an effect modifier of traffic on lung function in this study, which could be true or I did not have enough data to know. When evaluating environmental justice, children in areas with higher deprivation were located near

lower levels of traffic density. Neighborhood traffic density though, in this cohort of children with asthma, was causally associated with a reduction in lung function, when I compared the effect of high traffic density to lower traffic density. While neighborhood deprivation did not change the estimates of traffic on lung function, I found that there is likely to be a complex relation between deprivation factors that are not classically defined as poverty. Asthma can be worsened by the environment, as illustrated via traffic density, but it can also be improved by environmental factors. There is still work to be done in the arena of area-level effects on asthma; however, this dissertation contributed an important piece to our understanding of local traffic on lung function among children with asthma.

Recommendations

Future work ought to include further investigation on traffic and other measurements of lung function, such as the ratio of FEV_1 to FVC, FEF_{25-75}, and FEV_{75} to understand if the results were specific to FEV_1. This analysis could be extended to use the entire residential history of the FACES cohort. Each year a child spent in the study could be assigned a traffic exposure, instead of just the last year spent at the residence. Also, the IRT analysis could be expanded to use more measures of deprivation, and the IRT analysis could undergo more extensive model exploration with the addition of extra variables and fine tuning the IRT analysis techniques.

References

1. Bloom B, RA C. Summary Health Statistics for U.S. Children: National Health Interview Survey, 2006. Vital Health Stat 2007;10(234).
2. Akinbami L. The state of childhood asthma, United States, 1980-2005. Adv Data 2006(381):1-24.
3. CDC. Morbidity and Mortality Weekly Report. Vital Signs: Asthma Prevalence, Disease Characteristics, and Self-Management Education--United States, 2001-2009. Vol. 60, 2011;547-552.
4. CDC. National Hospital Ambulatory Medical Care Survey: 2008 Emergency Department Summary Tables. Ahead of Press, National Health Statistics Reports, 2011.
5. Corburn J, Osleeb J, Porter M. Urban asthma and the neighbourhood environment in New York City. Health Place 2006;12(2):167-79.
6. Litonjua AA, Carey VJ, Weiss ST, Gold DR. Race, socioeconomic factors, and area of residence are associated with asthma prevalence. Pediatr Pulmonol 1999;28(6):394-401.
7. Oakes JM. The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. Soc Sci Med 2004;58(10):1929-52.
8. Yen IH, Syme SL. The social environment and health: a discussion of the epidemiologic literature. Annu Rev Public Health 1999;20:287-308.
9. Pickett KE, Pearl M. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review. J Epidemiol Community Health 2001;55(2):111-22.
10. Diez Roux AV. Investigating neighborhood and area effects on health. Am J Public Health 2001;91(11):1783-9.
11. Huie SAB. The concept of neighborhood in health and mortality research. Sociological Spectrum 2001;21(3):341-358.
12. Poortinga W, Dunstan FD, Fone DL. Neighbourhood deprivation and self-rated health: The role of perceptions of the neighbourhood and of housing problems. Health Place 2007.
13. Ellen IG, Mijanovich T, Dillman KN. Neighborhood effects on health: Exploring the links and assessing the evidence. Journal of Urban Affairs 2001;23(3-4):391-408.
14. Cubbin C, Hadden WC, Winkleby MA. Neighborhood context and cardiovascular disease risk factors: the contribution of material deprivation. Ethn Dis 2001;11(4):687-700.
15. LeClere FB, Rogers RG, Peters K. Neighborhood social context and racial differences in women's heart disease mortality. J Health Soc Behav 1998;39(2):91-107.
16. Jacobs DE, Wilson J, Dixon SL, Smith J, Evens A. The relationship of housing and population health: a 30-year retrospective analysis. Environ Health Perspect 2009;117(4):597-604.
17. Riva M, Gauvin L, Barnett TA. Toward the next generation of research into small area effects on health: a synthesis of multilevel investigations published since July 1998. J Epidemiol Community Health 2007;61(10):853-61.
18. Chaix B. Geographic life environments and coronary heart disease: a literature review, theoretical contributions, methodological updates, and a research agenda. Annu Rev Public Health 2009;30:81-105.
19. Buekens P, Klebanoff M. Preterm birth research: from disillusion to the search for new mechanisms. Paediatr Perinat Epidemiol 2001;15 Suppl 2:159-61.

20. Grady S. Racial residential segregation impacts on low birth weight using improved neighborhood boundary definitions. Spatial and Spatio-temporal Epidemiology 2010;1:239-249.
21. Masi CM, Hawkley LC, Piotrowski ZH, Pickett KE. Neighborhood economic disadvantage, violent crime, group density, and pregnancy outcomes in a diverse, urban population. Soc Sci Med 2007;65(12):2440-57.
22. Messer LC, Kaufman JS, Dole N, Savitz DA, Laraia BA. Neighborhood crime, deprivation, and preterm birth. Ann Epidemiol 2006;16(6):455-62.
23. Janevic T, Stein CR, Savitz DA, Kaufman JS, Mason SM, Herring AH. Neighborhood Deprivation and Adverse Birth Outcomes among Diverse Ethnic Groups. Annals of Epidemiology 2010;20(6):445-451.
24. Nkansah-Amankra S. Neighborhood contextual factors, maternal smoking, and birth outcomes: multilevel analysis of the South Carolina PRAMS survey, 2000-2003. J Womens Health (Larchmt) 2010;19(8):1543-52.
25. Auger N, Daniel M, Platt RW, Wu Y, Luo ZC, Choiniere R. Association between perceived security of the neighbourhood and small-for-gestational-age birth. Paediatr Perinat Epidemiol 2008;22(5):467-77.
26. Ross CE, Mirowsky J. Neighborhood disadvantage, disorder, and health. Journal of Health and Social Behavior 2001;42(3):258-276.
27. Salmond C, Crampton P, Hales S, Lewis S, Pearce N. Asthma prevalence and deprivation: a small area analysis. J Epidemiol Community Health 1999;53(8):476-80.
28. Watson JP, Cowen P, Lewis RA. The relationship between asthma admission rates, routes of admission, and socioeconomic deprivation. Eur Respir J 1996;9(10):2087-93.
29. Kyle RG, Kukanova M, Campbell M, Wolfe I, Powell P, Callery P. Childhood disadvantage and emergency admission rates for common presentations in London: an exploratory analysis. Arch Dis Child 2011;96(3):221-6.
30. Sternthal MJ, Jun HJ, Earls F, Wright RJ. Community violence and urban childhood asthma: a multilevel analysis. Eur Respir J 2010;36(6):1400-9.
31. Clougherty JE, Levy JI, Kubzansky LD, Ryan PB, Suglia SF, Canner MJ, Wright RJ. Synergistic effects of traffic-related air pollution and exposure to violence on urban asthma etiology. Environ Health Perspect 2007;115(8):1140-6.
32. Wright RJ, Steinbach SF. Violence: an unrecognized environmental exposure that may contribute to greater asthma morbidity in high risk inner-city populations. Environ Health Perspect 2001;109(10):1085-9.
33. Swahn MH, Bossarte RM. The associations between victimization, feeling unsafe, and asthma episodes among US high-school students. Am J Public Health 2006;96(5):802-4.
34. Jeffrey J, Sternfeld I, Tager I. The association between childhood asthma and community violence, Los Angeles County, 2000. Public Health Rep 2006;121(6):720-8.
35. Wright RJ, Mitchell H, Visness CM, Cohen S, Stout J, Evans R, Gold DR. Community violence and asthma morbidity: the Inner-City Asthma Study. Am J Public Health 2004;94(4):625-32.
36. Byrd RS, Joad JP. Urban asthma. Curr Opin Pulm Med 2006;12(1):68-74.
37. Sandberg S, Paton JY, Ahola S, McCann DC, McGuinness D, Hillary CR, Oja H. The role of acute and chronic stress in asthma attacks in children. Lancet 2000;356(9234):982-7.

38. Wright RJ. Stress and atopic disorders. J Allergy Clin Immunol 2005;116(6):1301-6.
39. Lehrer P, Feldman J, Giardino N, Song HS, Schmaling K. Psychological aspects of asthma. J Consult Clin Psychol 2002;70(3):691-711.
40. Chen E, Hanson MD, Paterson LQ, Griffin MJ, Walker HA, Miller GE. Socioeconomic status and inflammatory processes in childhood asthma: the role of psychological stress. J Allergy Clin Immunol 2006;117(5):1014-20.
41. Jakeways N, McKeever T, Lewis SA, Weiss ST, Britton J. Relationship between FEV1 reduction and respiratory symptoms in the general population. Eur Respir J 2003;21(4):658-63.
42. Ashley F, Kannel WB, Sorlie PD, Masson R. Pulmonary function: relation to aging, cigarette habit, and mortality. Ann Intern Med 1975;82(6):739-45.
43. Ebi-Kryston KL. Respiratory symptoms and pulmonary function as predictors of 10-year mortality from respiratory disease, cardiovascular disease, and all causes in the Whitehall Study. J Clin Epidemiol 1988;41(3):251-60.
44. Lebecque P, Kiakulanda P, Coates AL. Spirometry in the asthmatic child: is FEF25-75 a more sensitive test than FEV1/FVC? Pediatr Pulmonol 1993;16(1):19-22.
45. Meliker JR, Sloan CD. Spatio-temporal epidemiology: principles and opportunities. Spatial and Spatio-temporal Epidemiology 2011;2:1-9.
46. Messer LC, Laraia BA, Kaufman JS, Eyster J, Holzman C, Culhane J, Elo I, Burke JG, O'Campo P. The development of a standardized neighborhood deprivation index. J Urban Health 2006;83(6):1041-62.
47. Drukker M, van Os J. Mediators of neighbourhood socioeconomic deprivation and quality of life. Soc Psychiatry Psychiatr Epidemiol 2003;38(12):698-706.
48. Kawachi I, Kennedy BP, Wilkinson RG. Crime: social disorganization and relative deprivation. Soc Sci Med 1999;48(6):719-31.
49. Hay GC, Whigham PA, Kypri K, Langley JD. Neighbourhood deprivation and access to alcohol outlets: a national study. Health Place 2009;15(4):1086-93.
50. Fisher JB, Kelly M, Romm J. Scales of environmental justice: combining GIS and spatial analysis for air toxics in West Oakland, California. Health Place 2006;12(4):701-14.
51. Marshall JD, Brauer M, Frank LD. Healthy neighborhoods: walkability and air pollution. Environ Health Perspect 2009;117(11):1752-9.
52. Greenberg M, Schneider D. Hazardous Waste Site Remediation, Neighborhood Change, and Neighborhood Quality. Environ Health Perspect 1994;102(6-7):542-7.
53. Coulton CJ, Korbin J, Chan T, Su M. Mapping residents' perceptions of neighborhood boundaries: a methodological note. Am J Community Psychol 2001;29(2):371-83.
54. Boyle MH, Willms JD. Place effects for areas defined by administrative boundaries. Am J Epidemiol 1999;149(6):577-85.
55. Haney WG, Knowles ES. Perception of Neighborhoods by City and Suburban Residents. Human Ecology 1978;6(2):201-214.
56. Chaskin RJ. Perspectives on neighborhood and community: A review of the literature. Social Service Review 1997;71(4):521-547.
57. Messer LC, Oakes JM, Mason S. Effects of socioeconomic and racial residential segregation on preterm birth: a cautionary tale of structural confounding. Am J Epidemiol 2010;171(6):664-73.

58. Helfand WH, Lazarus J, Theerman P. Donora, Pennsylvania: an environmental disaster of the 20th century. Am J Public Health 2001;91(4):553.
59. Davis DL, Bell ML, Fletcher T. A look back at the London smog of 1952 and the half century since. Environ Health Perspect 2002;110(12):A734-5.
60. Nemery B, Hoet PH, Nemmar A. The Meuse Valley fog of 1930: an air pollution disaster. Lancet 2001;357(9257):704-8.
61. Holgate ST. Air pollution and health. San Diego, Calif.: Academic Press, 1999.
62. Shy CM. Epidemiologic evidence and the United States air quality standards. Am J Epidemiol 1979;110(6):661-71.
63. Logan WP. Mortality in the London fog incident, 1952. Lancet 1953;1(6755):336-8.
64. Greenburg L, Jacobs MB, Drolette BM, Field F, Braverman MM. Report of an air pollution incident in New York City. November 1953. Public Health Rep 1962;77:7-16.
65. Dominici F, Peng RD, Barr CD, Bell ML. Protecting human health from air pollution: shifting from a single-pollutant to a multipollutant approach. Epidemiology 2010;21(2):187-94.
66. Rosenlund M, Berglind N, Pershagen G, Hallqvist J, Jonson T, Bellander T. Long-term exposure to urban air pollution and myocardial infarction. Epidemiology 2006;17(4):383-90.
67. Tonne C, Melly S, Mittleman M, Coull B, Goldberg R, Schwartz J. A case-control analysis of exposure to traffic and acute myocardial infarction. Environ Health Perspect 2007;115(1):53-7.
68. Kunzli N, Kaiser R, Medina S, Studnicka M, Chanel O, Filliger P, Herry M, Horak F, Jr., Puybonnieux-Texier V, Quenel P, Schneider J, Seethaler R, Vergnaud JC, Sommer H. Public-health impact of outdoor and traffic-related air pollution: a European assessment. Lancet 2000;356(9232):795-801.
69. Jerrett M, Finkelstein MM, Brook JR, Arain MA, Kanaroglou P, Stieb DM, Gilbert NL, Verma D, Finkelstein N, Chapman KR, Sears MR. A cohort study of traffic-related air pollution and mortality in Toronto, Ontario, Canada. Environ Health Perspect 2009;117(5):772-7.
70. Jerrett M, Burnett RT, Pope CA, 3rd, Ito K, Thurston G, Krewski D, Shi Y, Calle E, Thun M. Long-term ozone exposure and mortality. N Engl J Med 2009;360(11):1085-95.
71. Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects In: Pollution HPotHEoT-RA, ed. Boston, MA: Health Effects Institute, 2010.
72. Jerrett M, Shankardass K, Berhane K, Gauderman WJ, Kunzli N, Avol E, Gilliland F, Lurmann F, Molitor JN, Molitor JT, Thomas DC, Peters J, McConnell R. Traffic-related air pollution and asthma onset in children: a prospective cohort study with individual exposure measurement. Environ Health Perspect 2008;116(10):1433-8.
73. McConnell R, Berhane K, Gilliland F, London SJ, Islam T, Gauderman WJ, Avol E, Margolis HG, Peters JM. Asthma in exercising children exposed to ozone: a cohort study. Lancet 2002;359(9304):386-91.
74. Akinbami LJ, Lynch CD, Parker JD, Woodruff TJ. The association between childhood asthma prevalence and monitored air pollutants in metropolitan areas, United States, 2001-2004. Environ Res;110(3):294-301.

75. O'Connor GT, Neas L, Vaughn B, Kattan M, Mitchell H, Crain EF, Evans R, 3rd, Gruchalla R, Morgan W, Stout J, Adams GK, Lippmann M. Acute respiratory health effects of air pollution on children with asthma in US inner cities. J Allergy Clin Immunol 2008;121(5):1133-1139 e1.
76. Wilhelm M, Meng YY, Rull RP, English P, Balmes J, Ritz B. Environmental public health tracking of childhood asthma using California health interview survey, traffic, and outdoor air pollution data. Environ Health Perspect 2008;116(9):1254-60.
77. McConnell R, Berhane K, Yao L, Jerrett M, Lurmann F, Gilliland F, Kunzli N, Gauderman J, Avol E, Thomas D, Peters J. Traffic, susceptibility, and childhood asthma. Environ Health Perspect 2006;114(5):766-72.
78. Clark NA, Demers PA, Karr CJ, Koehoorn M, Lencar C, Tamburic L, Brauer M. Effect of early life exposure to air pollution on development of childhood asthma. Environ Health Perspect 2010;118(2):284-90.
79. U.S. Environmental Protection Agency, Office of Transportation and Air Quality. Greenhouse Gas Emissons from the U.S. Transportation Sector 1990-2003 Fairfax, V.A., 2006.
80. Pereira G, De Vos AJ, Cook A. Residential traffic exposure and children's emergency department presentation for asthma: a spatial study. Int J Health Geogr 2009;8:63.
81. Zwack LM, Paciorek CJ, Spengler JD, Levy JI. Modeling spatial patterns of traffic-related air pollutants in complex urban terrain. Environ Health Perspect 2011;119(6):852-9.
82. Pirkle JL, Kaufmann RB, Brody DJ, Hickman T, Gunter EW, Paschal DC. Exposure of the U.S. population to lead, 1991-1994. Environ Health Perspect 1998;106(11):745-50.
83. McEntee JC, Ogneva-Himmelberger Y. Diesel particulate matter, lung cancer, and asthma incidences along major traffic corridors in MA, USA: A GIS analysis. Health Place 2008;14(4):817-28.
84. Woodruff TJ, Parker JD, Kyle AD, Schoendorf KC. Disparities in exposure to air pollution during pregnancy. Environ Health Perspect 2003;111(7):942-6.
85. Mohai P, Lantz PM, Morenoff J, House JS, Mero RP. Racial and socioeconomic disparities in residential proximity to polluting industrial facilities: evidence from the Americans' Changing Lives Study. Am J Public Health 2009;99 Suppl 3:S649-56.
86. Bullard RD. Invisible Houston : the Black experience in boom and bust. Texas A & M southwestern studies no. 6. 1st ed. College Station: Texas A&M University Press, 1987.
87. Bullard RD. Dumping in Dixie : race, class, and environmental quality. Boulder: Westview Press, 1990.
88. Bullard RD. Environmental Justice for All: It's the Right Thing To Do. Journal of Environmental Law and Litigation 1994;9:291-308.
89. Clinton WJ. Federal Actions to Address Environmental Justice in Minority Populations and Low-Income Populations, Executive Order 12898. Federal Register, Vol. 59, No. 32, 1994;5.
90. U.S. GAO. Environmental Justice: EPA Needs to Take Additional Actions to Help Ensure Effective Implementation. Washington D.C.: U.S. Government Accountability Office, 2011.

91. Stuart AL, Mudhasakul S, Sriwatanapongse W. The social distribution of neighborhood-scale air pollution and monitoring protection. J Air Waste Manag Assoc 2009;59(5):591-602.
92. Melvin PM. Changing Contexts - Neighborhood Definition and Urban Organization. American Quarterly 1985;37(3):357-367.
93. Green HW, Truesdell LE. *Census Tracts in American Cities: A Brief History of the Census Tract Movement, with an Outline of Procedure and Suggested Modifications*. US Department of Commerce, Bureau of the Census, 1937.
94. Tatalovich Z, Wilson JP, Milam JE, Jerrett ML, McConnell R. Competing definitions of contextual environments. Int J Health Geogr 2006;5:55.
95. Openshaw S, Taylor P. The modifiable areal unit problem. In: Wrigley N, Bennett RJ, eds. Quantitative geography : a British view. London ; Boston: Routledge & Kegan Paul, 1981;60-69.
96. Openshaw S, Taylor P. A million or so correlation coefficients. In: Wrigley N, ed. Statistical applications in the spatial sciences. London: Pion, 1979;127-144.
97. Perry CA, Heydecker WD, Adams T. Neighborhood and community planning. Comprising three monographs: The neighborhood unit. New York,: Arno Press, 1974.
98. Gale S, Magzamen S, Radke J, Tager I. Crime, neighborhood deprivation, and asthma: a GIS approach to define and assess neighborhoods. Spatial and Spatio-temporal Epidemiology 2011;2:59-67.
99. Cornfield J. Randomization by group: a formal analysis. Am J Epidemiol 1978;108(2):100-2.
100. Poole C. On the origin of risk relativism. Epidemiology 2010;21(1):3-9.
101. Cornfield J, Haenszel W, Hammond EC, Lilienfeld AM, Shimkin MB, Wynder EL. Smoking and lung cancer: recent evidence and a discussion of some questions. J Natl Cancer Inst 1959;22(1):173-203.
102. Kaufman JS. Toward a more disproportionate epidemiology. Epidemiology 2010;21(1):1-2.
103. Maldonado G, Greenland S. Estimating causal effects. Int J Epidemiol 2002;31(2):422-9.
104. Mortimer KM, Fallot A, Balmes JR, Tager IB. Evaluating the use of a portable spirometer in a study of pediatric asthma. Chest 2003;123(6):1899-907.
105. Mortimer K, Neugebauer R, Lurmann F, Alcorn S, Balmes J, Tager I. Early-lifetime exposure to air pollution and allergic sensitization in children with asthma. J Asthma 2008;45(10):874-81.
106. Margolis HG, Mann JK, Lurmann FW, Mortimer KM, Balmes JR, Hammond SK, Tager IB. Altered pulmonary function in children with asthma associated with highway traffic near residence. Int J Environ Health Res 2009;19(2):139-55.
107. Mann JK, Balmes JR, Bruckner TA, Mortimer KM, Margolis HG, Pratt B, Hammond SK, Lurmann FW, Tager IB. Short-term effects of air pollution on wheeze in asthmatic children in Fresno, California. Environ Health Perspect 2010;118(10):1497-502.
108. Tager IB, Lurmann FW, Haight T, Alcorn S, Penfold B, Hammond SK. Temporal and spatial patterns of ambient endotoxin concentrations in Fresno, California. Environ Health Perspect 2010;118(10):1490-6.

109. Bigham JM, Rice TM, Pande S, Lee J, Park SH, Gutierrez N, Ragland DR. Geocoding police collision report data from California: a comprehensive approach. Int J Health Geogr 2009;8:72.
110. Jerrett M, McConnel R, Chang CCR, Wolch J, Reynolds K, Lurmann F, Gilliland F, Berhane K. Automobile traffic around the home and attained body mass index: A longitudinal cohort study of children aged 10-18 years. Preventive Medicine 2010;50:S50-S58.
111. Bailey TC, Gatrell AC. Interactive Spatial Data Analysis. Harlow: Longman, 1995.
112. Silverman BW. Density estimation for statistics and data analysis. Monographs on statistics and applied probability. London ; New York: Chapman and Hall, 1986.
113. Kan H, Heiss G, Rose KM, Whitsel EA, Lurmann F, London SJ. Prospective analysis of traffic exposure as a risk factor for incident coronary heart disease: the Atherosclerosis Risk in Communities (ARIC) study. Environ Health Perspect 2008;116(11):1463-8.
114. Zhu Y, Hinds WC, Kim S, Sioutas C. Concentration and size distribution of ultrafine particles near a major highway. J Air Waste Manag Assoc 2002;52(9):1032-42.
115. Zhou Y, Levy JI. Factors influencing the spatial extent of mobile source air pollution impacts: a meta-analysis. Bmc Public Health 2007;7.
116. Hu SS, Fruin S, Kozawa K, Mara S, Paulson SE, Winer AM. A wide area of air pollutant impact downwind of a freeway during pre-sunrise hours. Atmospheric Environment 2009;43(16):2541-2549.
117. Shi JP, Khan AA, Harrison RM. Measurements of ultrafine particle concentration and size distribution in the urban atmosphere. Science of the Total Environment 1999;235(1-3):51-64.
118. Su JG, Jerrett M, Beckerman B, Wilhelm M, Ghosh JK, Ritz B. Predicting traffic-related air pollution in Los Angeles using a distance decay regression selection strategy. Environ Res 2009;109(6):657-70.
119. Mitchell A. The ESRI Guide to GIS Analysis, Volume 2. ESRI Press, 2005.
120. Shriver K. Influence of Environmental Design on Pedestrian Travel Behavior in Four Austin Neighborhoods. Transportation Research Record 1997;1578(Paper 961076).
121. Donahue R. Pedestrians and Park Planning: How Far Will People Walk? The Trust for Public Land. San Francisco, 2011.
122. Pearl J. Causality: models, reasoning, and inference. 2nd ed. New York: Cambridge University Press, 2009.
123. Dudoit S, van der Laan M. Multiple testing procedures with applications to genomics. Springer series in statistics. New York ; London: Springer, 2008.
124. van der Laan MJ, Rubin D. Targeted maximum likelihood learning. The International Journal of Biostatistics 2006;2(1).
125. Gruber S, van der Laan MJ. Targeted maximum likelihood estimation: a gentle introduction. U.C. Berkeley Division of Biostatistics Working Paper Series 2009; Working Paper 252.
126. Sinisi SE, van der Laan MJ. Deletion/substitution/addition algorithm in learning with applications in genomics. Stat Appl Genet Mol Biol 2004;3:Article18.
127. Noth EM, Hammond SK, Biging GS, Tager IB. A spatial-temporal regression model to predict daily outdoor residential PAH concentrations in an epidemiologic study in Fresno, CA. Atmospheric Environment 2011;45(14):2394-2403.

128. Wilhelm M, Ghosh JK, Su J, Cockburn M, Jerrett M, Ritz B. Traffic-related air toxics and term low birth weight in Los Angeles County, California. Environ Health Perspect 2011;120(1):132-8.
129. US Census Bureau. A compass for understanding and using American Community Survey data: what the media need to know. Washington, DC: US Government Printing Office, 2008.
130. Tager IB, Balmes J, Lurmann F, Ngo L, Alcorn S, Kunzli N. Chronic exposure to ambient ozone and lung function in young adults. Epidemiology 2005;16(6):751-9.
131. Chilmonczyk BA, Salmun LM, Megathlin KN, Neveux LM, Palomaki GE, Knight GJ, Pulkkinen AJ, Haddow JE. Association between exposure to environmental tobacco smoke and exacerbations of asthma in children. N Engl J Med 1993;328(23):1665-9.
132. Wang X, Dockery DW, Wypij D, Fay ME, Ferris BG, Jr. Pulmonary function between 6 and 18 years of age. Pediatr Pulmonol 1993;15(2):75-88.
133. Schoenberg JB, Beck GJ, Bouhuys A. Growth and decay of pulmonary function in healthy blacks and whites. Respir Physiol 1978;33(3):367-93.
134. Dockery DW. Percentile curves for evaluation of repeated measures of lung function. Occup Med 1993;8(2):323-38.
135. Gale SL, Noth EM, Mann J, Balmes J, Hammond SK, Tager IB. Polycyclic aromatic hydrocarbon exposure and increased wheeze in a cohort of children with asthma in Fresno, CA. Journal of Exposure Science and Environmental Epidemiology 2012; Accepted for publication March 2012.
136. Cakmak S, Mahmud M, Grgicak-Mannion A, Dales RE. The influence of neighborhood traffic density on the respiratory health of elementary schoolchildren. Environ Int 2012;39(1):128-32.
137. Strickland MJ, Darrow LA, Klein M, Flanders WD, Sarnat JA, Waller LA, Sarnat SE, Mulholland JA, Tolbert PE. Short-term associations between ambient air pollutants and pediatric asthma emergency department visits. Am J Respir Crit Care Med 2010;182(3):307-16.
138. Cook AG, deVos AJ, Pereira G, Jardine A, Weinstein P. Use of a total traffic count metric to investigate the impact of roadways on asthma severity: a case-control study. Environ Health 2011;10:52.
139. Padula A, Mortimer K, Lurmann F, Hubbard A, Tager I. Exposure to traffic related air pollution during pregnancy and term low birth weight: an application of semi-parametric methods. In press at American Journal of Epidemiology 2012.