This TIE Fellowship report summarizes the methods employed and preliminary results and conclusions of a project to investigate whether C-reactive protein (CRP), a measure of systemic inflammation that has been associated with cardiovascular disease as well as several other chronic health conditions, is a meaningful biomarker of exposure to vehicle emissions for people living in close proximity to highways. In addition, the potential mediating effects of variation in the CRP gene and pre-existing health conditions were explored. Other measures of stress (hypertension, epinephrine, norepinephrine, salivary cortisol) and inflammation (albumin, homocysteine, white blood cell count) were also evaluated with respect to traffic exposure. The final budget account is included at the end.

Project Overview

Health status, demographic, and blood and genetic data were collected for 1020 older adults (50-70 years) by the Boston Puerto Rican Center for Population Health and Health Disparities Project (PRHDP) which resides in the Tufts School of Nutrition. My initial hypothesis addressed the associations between residential distance to highways and the inflammatory biomarker CRP. Though the study population resided in primarily urban areas, until I developed geographic coordinates for the case addresses, the number of cases residing within 200 meters of highways (the distance of concern supported by several previous studies), and whether this number would provide sufficient statistical power was uncertain.

This project was the first to verify PRHDP participant addresses, correct errors and then utilize this dataset to assess spatial influences. This laborious task determined that the number of cases living with 200 meters of highways was not sufficient to provide adequate power for a statistically significant evaluation. While not near highways in sufficient numbers, study participants clearly resided in the urban cores of Boston, Chelsea and Lawrence. I hypothesized that these areas also presented a substantial traffic burden, yet the literature was unclear on whether a dose-response relationship could be ascertained for levels of traffic exposure lower than those found near highways and whether a valid method of exposure characterization could be developed to address the heterogeneity within my study areas. With much invested, and in close consultation with my advisors, I decided to reframe my research question to test various methods for characterizing traffic exposures at these lower yet more pervasive levels and to use the available health data to explore the utility and consistency of these methods. I have been invited to present a comparison of these methods at the International Society for Environmental Epidemiology conference in October, 2008. I expect to complete manuscripts on this component and the health assessment component in fulfillment of my dissertation requirement by the end of 2008. A brief synopsis of my study approach and preliminary research results is included below.
Study Approach

My specific aim was to assess the spatial variation of disease, particularly inflammation and stress, with respect to traffic density and proximity controlling for pre-existing conditions and other socioeconomic variables. My research included the following components:

1. Verify case addresses and using GIS tools, derive the latitude and longitude, or XY coordinates, for each of these residential locations – a process called geocoding.

2. Select the relevant descriptive, lifestyle, socioeconomic, health-status, and biomarker variables from among the 1300 variables collected by the PRHDP to assess traffic exposures and health. Identify predictor and outcome variables along with potential confounders and effect modifiers.

3. Collect and synthesize available traffic volume and emissions data, models, and other information for the study area. Develop a set of traffic exposure variables to distinguish different levels of traffic burden for individual residential locations.

4. Assess associations between traffic burden and the selected health status variables and biomarkers.

5. Compare the methods of exposure characterization for consistency, relative strengths and limitations.

The first two components generally followed established spatial analysis and epidemiologic study protocols modified when necessary to address unique study features. For example, it was necessary to establish a geocoding hierarchy involving three methods to obtain coordinates: parcel matching, street network matching, and Google Earth/Google maps. Parcel maps were only available for Boston, Chelsea, and Cambridge and in these cities, public housing sites were not well represented. An estimated 25 percent of study cases reside in public housing. For cases outside these areas, or not matched to parcel maps a second method of geocoding was conducted using ArcGIS StreetMap USA, a nationwide street network for map visualization, geocoding, and routing available as part of the ArcGIS software. Scores were derived from the first two methods and each “match” was reviewed to determine the reason for scores less than 100 (on a scale of 0-100). Typical reasons for scores between 80-99 included slight misspellings, or the use of the abbreviation for street or road (st., rd). Scores below 80 were checked against Google Earth which generally resulted in identification of zip code errors or clarification of neighborhood and street names. Complicating the positional confirmation process, in Boston’s inner core, there are 18 distinct neighborhoods with over 200 instances of the same road name being used for two or more different non-contiguous roads. Together with an extensive address verification procedure, the goal of geocoding was to minimize positional error in deriving the address coordinates, given the focus on the effects of traffic at a localized scale.
The study population includes primarily females (72 percent) living in apartment complexes or multi-family units (70 percent) who are not currently working (70 percent), have household incomes below 120% of the poverty threshold (66 percent) and a median educational level of 9th grade. A majority (52 percent) are obese (body mass index over 30), have hypertension (70 percent) or depression (57 percent) and 41 percent have diabetes. Nearly 25 percent have diabetes, hypertension, and obesity. These characteristics, along with information on lifestyle (smoking, alcohol consumption), length of time at the present residence, and medications, were all considered in evaluating potential confounding factors (associated with both the exposure and the disease outcome but not in the causal pathway) and effect modifiers that might interact with the independent variables and influence the disease outcomes.

A three-tiered approach to characterizing traffic exposures was developed. As a mobile pollution source, traffic is an inherently dynamic phenomenon and no clear consensus has emerged for assessing the burden of traffic on health. Inner-city roadway networks are complex and heterogeneous. Tree-lined streets and quiet neighborhoods can quickly transition to densely travelled roads, congested intersections and areas with high amounts of automobile, truck and bus traffic over large portions of the day. Adding to this complexity are urban canyons, or streets cutting through blocks of buildings in a manner that influences wind speed, wind speed, and ultimately air quality. My goal was to develop an approach for characterizing exposure that was systematic, reproducible, transferable to other locales, yet captured the heterogeneity of my study areas. I surveyed available sources of data on traffic volumes and emissions as well as accepted and sanctioned models that predicted the same. Ultimately, I combined established tools from transportation planning with traffic monitoring data, and GIS spatial density analyses to assess traffic exposures in three different ways:

1. Traffic Analysis Zone (TAZ) assessment
2. Spatial Density Analysis using TAZ data
3. Proximity Assessments

A transportation analysis zone (TAZ) is the unit of geography most commonly used in conventional transportation planning models. The Central Transportation and Planning Staff (CTPS) at the Boston Region Metropolitan Planning Organization (MPO) are responsible for developing and modeling traffic volumes and emissions associated with each of 2727 TAZs in Massachusetts including the following parameters:

- Average daily vehicle miles travelled (VMT)
- Average daily vehicle hours travelled (VHT)
- Traffic-related emissions of carbon monoxide, nitrogen oxides, and volatile organic chemicals (CO, NOX, VOC)

Emissions were estimated using the EPA MOBILE6 model, the standard for transportation planning across the country. Each of these parameters was normalized based on the size of the
TAZ yielding values for vehicle miles travelled/square mile, vehicle hours travelled/square mile, etc. The TAZ in which each study case resided was determined based on their geographic coordinates and health outcomes were evaluated against their respective levels of traffic exposure, adjusting for confounders and effect modifiers.

Study cases were found to reside in 227 different TAZs. It was observed that a number of TAZs were very small (less than 0.1 square miles). TAZ are defined based on land use and demographic characteristics to maximize homogeneity and recognize physical barriers. In the process of demarcating TAZs, high traffic roads may often form TAZ boundaries with the traffic levels associated with that road split among the two adjacent TAZ. In some cases the traffic levels from a major road are counted within only one TAZ and the adjacent TAZ appears to have much lower traffic impacts. Residents close to the TAZ boundary may in fact be subject to the traffic effects of the adjacent TAZ, depending on road geography, meteorology, building design, etc. A spatial density analysis was conducted to examine the degree to which small TAZs may in fact be influenced by the traffic levels of their neighboring TAZ.

Spatial density analysis accounts for the size of a TAZ with respect to other TAZ and calculates the density of traffic levels in the vicinity smoothing the variation between them. A running weighted average of traffic levels within cells of a specified size is calculated over a specified radius of influence. Averages are calculated across the TAZ boundaries to account for substantial differences in traffic levels. I chose cells of 10 meters in diameter (also testing 25 and 50 meters) and looked out over a distance of 1000 meters (also testing 500 meters). Once a density map was derived, cases were assigned to one of five density levels that visually corresponded very closely with major highways and interchanges over the study area. Separate variables were created for DenseCode 5 as well as DenseCode 2, where all cases in density level 2 or higher were grouped into category 2, and DenseCode3, where all cases in density levels 3 or higher were grouped into category 3. Approximately 75 percent of cases fell into density level 1; 20 percent in density level 2 and the remaining 5 percent in density levels 3-5. Health outcomes were evaluated against traffic density levels, adjusting for confounders and effect modifiers.

The proximity assessment included developing several exposure variables to represent case proximity to roads with traffic volumes over 20,000 vehicles per day. A combination of data from Massachusetts Highway Department and CTPS was utilized to identify roads of concern, including a GIS compatible file with traffic count station positions and measurement data from 1997 – 2006. The final list of roads included approximate 40 major roads in various neighborhoods of Boston, including the Jamaicaicaway, Columbus Avenue, and Dorchester Avenue, and 10 highways and major routes such as Massachusetts Turnpike, Route 93, 28, 1, 99, 2, and 3 that pass through several cities and towns. Both 100 and 200 meter buffers around these roads and highways were constructed and cases residing within one or more of these buffers were identified. Separate variable was created for cases residing within 100 meters (Prox100), 200 meters (Prox200) and multiple buffers (ProxMult). Approximately 60 percent of cases did not live within any of the specified 100 meter or 200 meter buffers; 36 percent lived within one buffer; 4 percent lived within 2 buffers; and just over 1 percent (12 cases) lived within 3 or 4 buffers. Health outcomes were evaluated against road proximity levels, adjusting
for confounders and effect modifiers. Figure 1 shows a portion of the study area in downtown Boston with traffic density areas depicted in shades of blue, roadways of concern with buffers in light and dark pink, and study participants in small blue and red circles.

Figure 1

Preliminary Results

CRP was not found to be associated with traffic exposure as characterized by any of the traffic variables. One of the genetic markers for the CRP gene was statistically significantly associated (p<0.05) with CRP levels, consistent with the literature, but no interaction with traffic was observed in the CRP response while controlling for medications, the income/poverty ratio, alcohol, years living at this residence, and depression score. Several possible explanations exist for this lack of association, the most likely being that the selection criteria did not include traffic considerations but were focused on finding older Puerto Ricans, wherever they resided within the general study area. The background health status of this population was defined by elevated levels of diabetes, obesity and hypertension, all associated with elevated CRP levels. The median CRP level for this population was 3.7 mg/l, above the high risk level for cardiovascular disease of 3.0 mg/l specified by the American Heart Association. Statistically, it is postulated that there
were not enough cases across a range of traffic exposures with both high and low CRP levels to be able to discern a traffic-related impact.

Albumin, another inflammatory biomarker, was significantly associated with several of the traffic exposure variables including the two-level traffic density code, and all of the proximity variables (Prox100, Prox200, ProxMult, ProxMultCode) all at p<0.05 levels. Stratifying into cases that had all three relevant co-morbidities (diabetes, obesity, and hypertension), two of the co-morbidities, one or none resulted in only the groups with one or none remaining statistically significant (p=0.20 and 0.028) with coefficients of -0.044 and -0.04, respectively.

Hypertension was represented by a binary variable (0=systolic blood pressure less than 140 and diastolic blood pressure less than 90; 1 = above either the cutoff for systolic or diastolic or taking hypertension medications), and three continuous variables: (1) average (3 readings) systolic blood pressure; (2) average diastolic blood pressure measurements, and (3) average pulse pressure (systolic minus diastolic). Controlling for years lived at residence, and smoking, the two-level and three-level density codes were associated with the binary hypertension variable (p<0.05) with a coefficients of 1.8 (p=0.001) and 2.0 (p=0.002) representing odds ratios derived from logistic regression. Interestingly, the proximity traffic variables were not statistically significantly associated with the binary hypertension variable. Still with the binary hypertension variable, stratifying for smokers, ex-smokers, and non-smokers, the odds ratio for non-smokers (2.8 and 2.2 for the two-level and three-level density codes) were the highest and higher than the non-stratified odds ratio.

Neither average systolic nor diastolic blood pressure were associated with any of the traffic variables, however pulse pressure was significantly associated with all the proximity variables (p<0.05). Coefficients ranged from 2.9 for Prox100 to 1.4 for Prox200. When stratified, ex-smokers had the highest coefficients (3.0), remaining significant at the p<0.05 level. Stratifying for various potential confounders proved relevant: working/non-working yielded higher coefficients for workers (p<0.05); co-morbidities yielded a hierarchy of statistically significant coefficients (p<0.05) with cases with all three co-morbidities highest, followed by cases with two diseases and then one disease. Associations between pulse pressure and the proximity variables were not significant for cases with none of these diseases. Pulse pressure and hypertension were stronger outcomes for cases living less than 2 years at the present location.

Preliminary Conclusions

It appears that hypertension as represented by the binary variable is associated with traffic density and pulse pressure is associated with proximity to high traffic roads over 20,000 vehicles/day. The question of why hypertension is not associated with the traffic proximity variables and pulse pressure is not associated with the traffic density variables remains to be explored in my continuing work.
Albumin, a biomarker with emerging evidence of its association with stroke and cardiovascular outcomes, also appears to be associated with both traffic density and proximity for people without evidence of the co-morbidities of diabetes, hypertension and obesity.

Additional work will be conducted in several areas including the following: (1) the relative strengths and limitations of the various traffic exposure characterization methods including whether the same cases were defined as low, medium, and high traffic exposure across the different methods; (2) the role and inter-relationships of socioeconomic variables specifically the number of years lived at the present address, a potential measure of instability which is negatively correlated with multiple measures of traffic intensity; (3) the challenge of discerning traffic stress impacts in light of co-morbidities; and (4) preparation of manuscripts including background, introduction and discussion sections with complete reviews of literature.

BUDGET for $6000 TIE GRANT

- SPSS Software: $224
- GIS Intern for assistance in initial geocoding effort: $1080
- GIS Manual: $95
- Tufts health insurance: $1800
- Stipend during research, data assessment, draft preparation: $3200