Social Demographic Variables Obtained from Secondary Data Sources Relate to Spatial Trends in Traffic Volume Variables Obtained from PennDOT Data Sources

Mena Tausner* and Blanca E. Himes

Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA

**Correspondence can be sent to mtausner@sas.upenn.edu*

DEPARTMENT of BIOSTATISTICS EPIDEMIOLOGY & INFORMATICS

Introduction

Vehicular traffic, and the transportation sector in general, plays a significant factor in elevating the amount of fine particulate matter ($PM_{2.5}$) in surrounding areas with heightened traffic exposure due to resuspended dust created on roadways that remains in the air with the consistent motion of oncoming vehicles.[1] $PM_{2.5}$ is an easily-inhalable air pollutant commonly linked to pulmonary health conditions including asthma, COPD, and overall decreased respiratory function. Historically, there have been marked health disparities caused by such environmental factors like $PM_{2.5}$ in marginalized communities based on race/ethnicity and socioeconomic conditions.[2] The objective of this study was to determine whether traffic volume was higher in Pennsylvania census tracts with greater proportions of racial/ethnic minority residents and according to median family income as well as Environmental Justice Area status.

Linking PennDOT-Derived Traffic Data to Geospatially Varying Census-Derived Data

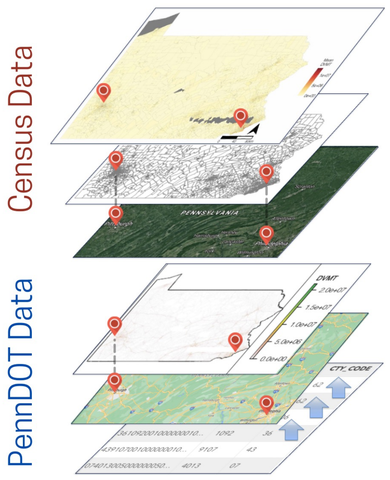

Figure 1. Raw by-roadway traffic volume data derived from PennDOT can be used to spatially link traffic exposure to social demographic environmental data on the 3218 geospatially-varying census tracts throughout Pennsylvania. In each map, Pennsylvania's urban regions of Philadelphia and Pittsburgh remain consistently marked for reference.

High resolution, by-roadway traffic volume data from the Pennsylvania Department of Transportation (PennDOT), all of which is publicly available, can be linked to geospatial US Census data through these general steps:

1. **Identify and obtain raw traffic volume data from PennDOT.** *PennDOT regularly collects traffic volume information from roadways across Pennsylvania for government road-planning, construction, and funding purposes.*
2. **Create a raster layer visualizing the raw traffic volume data with a gradient color scale across the Pennsylvania roadmap.** *Using the raster package within R Studio for geospatial data analysis, utilize the Average Amount of Daily Traffic (AADT) from PennDOT to calculate and plot the Daily Vehicle Miles of Travel (DVMT) for roadways across Pennsylvania based on road segment length.*
3. **Select external sources of geospatially varying data.** *Rich sources of geospatial data on the social environment of Pennsylvania, separated into the state's 3218 census tract boundaries, exist in the public domain via American Community Survey (ACS) data reported annually.*
4. **Integrate PennDOT-derived data to external data sources via a geospatial link.** *Traffic volume raster data can be linked to geospatial Census data using the extract function within the raster package in R Studio to calculate the mean value of the DVMT raster layer for each Pennsylvania census tract, identified by its respective obtained multipolygon geospatial coordinates.*
5. **Assess associations between traffic volume to geospatial social demographic trends derived for each Census tract.** *External geospatial data variables have their census tract-based spatial autocorrelation modeled explicitly.*

Methods

Study population Subjects were residents in Pennsylvania's 3218 census tracts.

Data Analysis software The analysis of all external data sources was conducted in R Studio using R Markdown files, taking advantage of R's capabilities in filtering raw data and intuitively plotting maps of geospatial data. The main R packages used throughout the study included sp, rgdal, sf, tidyverse, tidycensus, ggsn, mapview, RColorBrewer, viridis, tidyr, raster, rgeos, dplyr, scales, and broom.

Variables obtained from secondary sources The following external variables were linked to Census data throughout Pennsylvania via tract GEOID codes and spatial polygon coordinates applicable to the geospatial attributes of secondary data sources:

- **Traffic Volume:** We estimated each census tract's exposure to traffic by assessing the Current AADT (CUR_AADT) variable for roadways within the PennDOT raw traffic volume data and creating a 250-m circular buffer of each respective roadway segment to create a raster layer based on the calculated DVMT values for the immediate areas surrounding such roadways; we then extracted the mean DVMT for each census tract within the finalized traffic volume raster layer.
- **Racial/Ethnic Minority Proportion**: We determined the percentage of residents living within each census tract who identified their race as white and non-white (including Black or African American, American Indian or Alaskan Native, Asian, Native Hawaiian or Other Pacific Islander, Other, Two or More Races, etc.) based on variables extracted from ACS 5-year estimates for 2014-2019.
- **Median Family Income:** We determined the median family income of residents living within each census tract based on variables extracted from ACS 5-year estimates for 2014-2019.
- **Environmental Justice Area (EJA) Status:** We determined which census tracts qualified as Environmental Justice Areas (EJAs), defined as "any census tract where 20 percent or more individuals live at or below the federal poverty line, and/or 30 percent or more of the population identifies as a non-white minority", based on the Pennsylvania Department of Environmental Protection's EJA Census Block Group data for 2015.

Statistical analysis We used linear regression models to identify the relationship between social factors of tract-level median family income and percentage of white residents, as compared to racial/ethnic minority residents, associated with mean DVMT. We further used violin plot models to identify these social demographic factors associated with exacerbated traffic volume after dividing the study population into four groups for each factor based on quartile cut-off values determined by summarizing ACS data regarding tract-level median family income and percentage of white residents, respectively. A violin plot model was also used to identify EJA status factors associated with mean DVMT after dividing the study population into two groups of qualifying and non-qualifying EJA tracts.

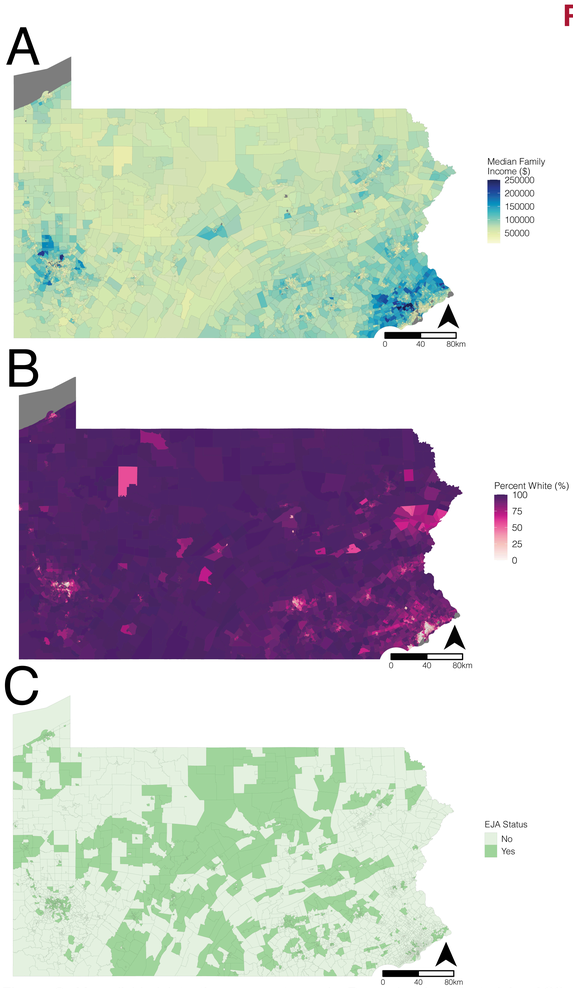

Figure 2. Map divided into the census tracts in Pennsylvania adjusted for ACS-derived population median family income only (A), adjusted for ACS-derived population percent white race only (B), and adjusted for Pennsylvania Department of Environmental Protection-derived EJA status (C). Regions of elevated quantities in the continuous variables (A and B) are shown in darker colors and lesser quantities in lighter colors on the respective gradient scales. Regions identified as EJAs (C) are shown in dark green, and regions not identified as EJAs are shown in light green.

Results

Figure 3. Scatter plots and corresponding linear regression model equations showing the association between tract-level mean DVMT versus ACS-derived median family income only (D) and versus ACS-derived population percent white race only (E) for all Pennsylvania tracts with available finite data values.

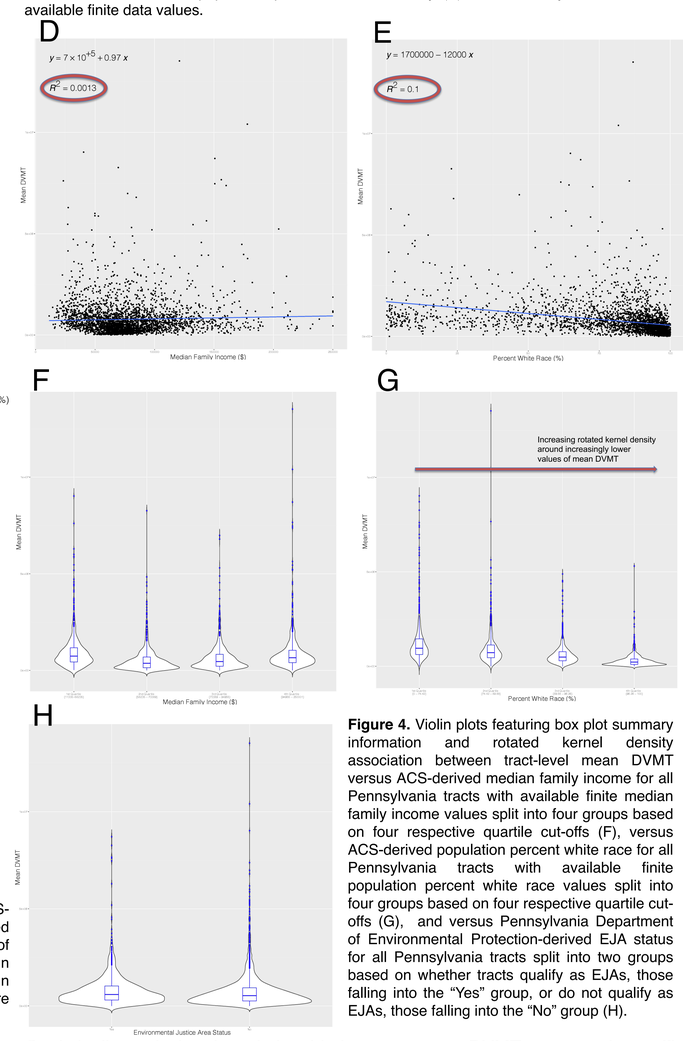

Figure 4. Violin plots featuring box plot summary information and rotated kernel density association between tract-level mean DVMT versus ACS-derived median family income for all Pennsylvania tracts with available finite median family income values split into four groups based on four respective quartile cut-offs (F), versus ACS-derived population percent white race for all Pennsylvania tracts with available finite population percent white race values split into four groups based on four respective quartile cut-offs (G), and versus Pennsylvania Department of Environmental Protection-derived EJA status for all Pennsylvania tracts split into two groups based on whether tracts qualify as EJAs, those falling into the "Yes" group, or do not qualify as EJAs, those falling into the "No" group (H).

Social demographic factors associated with exacerbated traffic volume Statistically analyzing the relationship between mean DVMT, representing traffic volume, versus median family income, proportion of white residents, and EJA status for all 3218 census tracts within Pennsylvania revealed that traffic volume was most consistently decreased for populations with heightened proportions of white residents as compared to other social factors of interest. As the percentage of racial/ethnic minority residents of census tracts increased, the violin plot's rotated kernel density showed that a heightened mean DVMT value became more probable within tracts constituting such lower-quartile percent white groups (Figure 4.G) as compared to the widely varying rotated kernel density among the four quartile-based groups in the violin plot assessing the association between traffic exposure and median family income (Figure 4.F). Linear regression model plots further demonstrate the closer association between traffic volume and proportion of white versus non-white residents with the increased coefficient of determination value $R^2 = 0.1$ (Figure 3.E) by nearly a factor of 10 compared to that of traffic volume and median family income where $R^2 = 0.0013$ (Figure 3.D). The violin plot assessing traffic volume compared to the two groups of EJA-qualifying versus non-qualifying tracts demonstrated slightly increased rotated kernel density at heightened mean DVMT values for EJAs, yet the box plot summaries prove quite similar in terms of median, 1st, and 3rd quartile mean DVMT values for both the "Yes" and "No" EJA groups (Figure 3.H).

Conclusion

Spatial trends in traffic volume rates in Pennsylvania determined from PennDOT data can be explained by differential patterns of census tract-level social environment factors, specifically proportion of racial/ethnic minority residents. These findings align with historical evidence for intensified traffic exposure and major highway placement as examples of unjust environmental racism across the US.[3] For future research, we plan to adjust our analysis for conditions of "urbanicity" to better assess by-income traffic volume exacerbation in cities versus rural areas across Pennsylvania.

Bibliography 1. Askariyeh et al. *Near-Road Traffic-Related Air Pollution: Resuspended PM 2.5 from Highways and Arterials* 2020; 2. Tessum et al. *$PM_{2.5}$ polluters disproportionately and systemically affect people of color in the United States* 2021; 3. Blakemore et al. *Interstate highways were touted as modern marvels. Racial injustice was part of the plan* 2021.

Funding This study was supported by the Penn Undergraduate Research Mentoring Program (PURM) as part of the University of Pennsylvania's Center for Undergraduate Research and Fellowships (CURF).