SUPPLEMENTAL MATERIAL

Title: Synergism of Short-Term Air Pollution Exposures and Neighborhood Disadvantage on Initial Stroke Severity

Authors: Jeffrey J. Wing PhD[1], Brisa N. Sánchez PhD[2], Sara D. Adar ScD[1], William J. Meurer MD[3,4], Lewis B. Morgenstern MD[1,3], Melinda A. Smith DrPH[3], Lynda D. Lisabeth PhD[1,3]

1. Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA

2. Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI, USA

3. Department of Neurology, Stroke Program, University of Michigan, Ann Arbor, MI, USA

4. Department of Emergency Medicine, University of Michigan, Ann Arbor, MI, USA

METHODS

Neighborhood disadvantage score

Census tracts were used as proxies for neighborhoods and stroke cases were assigned to the census track where they resided at the time of stroke. We identified 12 neighborhood-level Census variables that reflected sociodemographic domains of race/ethnicity and wealth/income from the 2000 Census, American Community Survey 2005-2009 and American Community Survey 2007-2011 including: percent non-Hispanic white, Hispanic, with more than high school education, with more than a bachelor degree, below poverty level,[1] unemployed, with managerial occupation, with income at least $50,000, with interest, dividend, or rental income, with household with no telephone, with household with no vehicle maintained, and median household income. Following the work of previous studies,[2] we created a composite score for neighborhood disadvantage based on z-scores for the individual neighborhood-level Census variables; variables were reverse coded as necessary, such that positive z-scores would represent higher neighborhood disadvantage. These scores were then applied to individual stroke subjects based on census tract of residence and year of stroke presentation. Average neighborhood disadvantage score was used when there were overlapping years of the American Community Survey data.

Air pollution and meteorology data

Hourly $PM_{2.5}$, O_3, and meteorological data from 2000-2012 were obtained from the Texas Commission on Environmental Quality's Texas Air Monitoring Information System for a centrally located monitor within the urban population. The use of a single fixed monitor may be a source measurement error if the daily fluctuations were not homogeneous across the county. However, another monitor in the county that measured 24-hour $PM_{2.5}$ every three to six days showed large correlations ($\rho \geq 80\%$) with the monitor used for our data collection. The topography of Nueces County is relatively flat, with elevation varying from sea level to a maximum elevation of 180 feet,[3] which suggests that air pollution exposures are likely to be highly correlated across space. The pollution levels in Nueces County are relatively low in comparison to other cities. For example, Los Angeles County has an annual average $PM_{2.5}$ of 17.2 ug/m^3 in 2008 and an annual average ozone concentration of 66.7 ppb. Relative humidity was calculated using dew point temperature from the Corpus Christi International airport with data from the National Oceanic and Atmospheric Administration. $PM_{2.5}$, ambient temperature, and relative humidity were averaged daily and the maximal 8-hour daily average was calculated for O_3 based on the timing for the US Environmental Protection Agency National Ambient Air Quality Standards. We focused on same-day $PM_{2.5}$ and previous-day O_3 concentrations based on previous work in this population exploring the association between short-term exposures and ischemic stroke risk.[4]

Statistical methods

Characteristics of the study population were summarized with descriptive statistics both overall and by neighborhood disadvantage score, comparing the 90th percentile to the 10th percentile. Associations between neighborhood disadvantage, air pollutants ($PM_{2.5}$ and O_3), and initial NIHSS were modeled in two ways based on previous work.[5] We *a priori* chose to analyze the data both continuously and dichotomously to aid in the interpretation of our findings. We used linear regression models for continuous NIHSS score and logistic regression models comparing the upper quartile of severity (severe stroke, NIHSS ≥7) to all lower quartiles (mild stroke, NIHSS <7).[6] Generalized estimating equations were used to account for clustering of stroke subjects within census tracts.[7] All models were adjusted for demographics (race/ethnicity, age and gender) and stroke risk factors (atrial fibrillation, coronary artery disease, diabetes, hypertension, and smoking status). The air pollution models were additionally adjusted for ambient temperature, humidity, day of the week, season, and a natural spline for time with 96 degrees of freedom (approximately 8 degrees of freedom per year). We assessed the functional forms of the neighborhood disadvantage, air pollutants and meteorological variables using quartiles and penalized polynomial splines. Although the distribution of NIHSS score was right skewed, its relation with air pollution was approximately linear and thus we maintained the original variable for ease of interpretation. We additionally found linear functional forms to be appropriate for neighborhood disadvantage and both meteorological variables.

Modeling was done in three stages. First, the main effects of neighborhood disadvantage and each air pollutant on NIHSS score were examined separately in single exposure models. Second, the main effects of neighborhood disadvantage and each air pollutant were examined together in dual exposure models. Third, the interaction between each air pollutant and neighborhood disadvantage was added to the dual exposure models. Additionally, multi-pollutant models with both air pollutants included were examined to explore the potential confounding effects.

Associations between neighborhood disadvantage and NIHSS score were summarized as a comparison of the 90^{th} to the 10^{th} percentile of the neighborhood disadvantage score. Differences in NIHSS score were calculated for $10\mu g/m^3$ or 10ppb change in $PM_{2.5}$ and O_3, respectively. Presence of effect modification of the air pollution-stroke severity association by neighborhood disadvantage was indicated by a statistical interaction term p-value<0.10. Estimates of the air pollution-stroke severity association at the 10^{th} and 90^{th} percentiles of neighborhood disadvantage score were reported for ease of interpretation of the effect modification.

All analyses were conducted using SAS version 9.3 (SAS Institute Inc, Cary, NC) and the R statistical package, version 3.0.1 (R Foundation for Statistical Computing, Vienna, Austria). The BASIC project was approved by the University of Michigan Institutional Review Board and each of the Nueces County hospital systems.

REFERENCES

1. Proctor BD, Semega JL, Kollar MA; U.S. Census Bureau. Income and Poverty in the United States : 2015. https://www.census.gov/content/dam/Census/library/publications/2016/demo/p60-256.pdf. Published September, 2016. Accessed July 11, 2017.

2. Lisabeth LD, Diez Roux AV, Escobar JD, Smith MA, Morgenstern LB. Neighborhood environment and risk of ischemic stroke: the brain attack surveillance in Corpus Christi (BASIC) Project. *Am. J. Epidemiol.* 2007;165:279–87.

3. National Elevation Dataset (NED) 2013. Texas Natural Resources Information System. https://tnris.org/data-catalog/entry/national-elevation-dataset-ned-2013/. Accessed June 27, 2013.

4. Wing JJ, Adar SD, Sánchez BN, Morgenstern LB, Smith MA, Lisabeth LD. Ethnic differences in ambient air pollution and risk of acute ischemic stroke. *Environ. Res.* 2015;143:62–7.

5. Kleindorfer D, Lindsell C, Alwell KA, Moomaw CJ, Woo D, Flaherty ML, et al. Patients living in impoverished areas have more severe ischemic strokes. *Stroke.* 2012;43:2055–9.

6. Vittinghoff E, Glidden D V., Shiboski SC, McCulloch CE. Regression Methods in Biostatistics. 2nd ed. New York: Springer; 2011.

7. Fitzmaurice GM, Laird NM, Ware JH. Applied Longitudinal Analysis. Hoboken, NJ: John Wiley & Sons, Inc.; 2011.