Examining Spatially Varying Relationships between Preterm Births and Ambient Air Pollution in Georgia using Geographically Weighted Logistic Regression

Jun Tu

Department of Geography and Anthropology
Kennesaw State University
Introduction

- Preterm Birth (PTB, birth occurring before the 37th week pregnancy) can affect the subsequent health status of individuals, including elevated mortality and morbidity in childhood, and children’s developmental disabilities.

- In 2013, about 36% of the infant deaths in US were due to PTB-related causes. PTB was responsible for 15% of the children death before 5 years of age globally.

- Thus, studies on the causes of PTB are extremely important to improve birth outcomes and human health.

- Numerous studies have found that PTB is associated with many environmental and socio-economic factors, such as air pollution, maternal smoking and drinking, prenatal care, education, races, and income.
• Air pollution is one of the most frequently studied factors

• The results are not consistent and the causal relationships have not been built.

• Non-significant, significant positive, and even significant negative relationships between PTB risk and prenatal exposures to air pollutants were found in different studies.

• The physical and social environment and human behaviors are changing over space.

• We assume the relationships between PTB and air pollutants change over space.
Logistic Regression is a commonly used statistical method to analyze the relationships

- Dependent variables: Preterm Birth (Yes/No)
- Independent variables: Individual-level, community-level socioeconomic, environmental, and behavioral factors

Conventional logistic regression is Global Statistics, Relationships are for the whole study area, and represent the average situation, assuming the relationships are constant over space

- May hide some local relationships and local causes of PTB
GWLR (Geographically Weighted Logistic Regression)

- An extension of the traditional standard logistic regression

\[
y_i = \frac{e^{(\beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki})}}{1 + e^{(\beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki})}}
\]

(1)

the mathematical expression of its geographically weighted version is:

\[
y_{(u_i, v_i)} = \frac{e^{(\beta_0(u_i, v_i) + \beta_1 (u_i, v_i)x_{1i} + \cdots + \beta_k (u_{ki}, v_{ki})x_{ki})}}{1 + e^{(\beta_0(u_i, v_i) + \beta_1 (u_i, v_i)x_{1i} + \cdots + \beta_k (u_{ki}, v_{ki})x_{ki})}}
\]

(2)

where \((u_i, v_i)\) are the location coordinates in space of point \(i\).

- Capture spatial variations by allowing regression model parameters to change over space

- Produce local regression results including local parameter estimates and the local \(R^2\) values for each regression point
In this study

- GWLR technique is used to examine the spatial variations in the relationships between the occurrence of PTB and concentrations of O3 and PM2.5 in Georgia.

Objectives

- To compare the associations of PTB with concentrations of O3 and PM 2.5 produced by global logistic regression and GWLR.
- To explore how the association of PTB with each air pollutant generated by GWLR vary spatially.
- How the spatially varying association of PTB with each air pollutant is affected by the socioeconomic and urbanization characteristics of communities.
Data Sources and Methods

Variables

• Dependent Variable: (binary)
  – Individual PTB (yes/no)

• Independent Variables:
  – Nine Individual-level birth and maternal factors
    • Infant Gender (Male, Female)
    • Parity (The order of the birth of the mother)
    • Mothers Race (White, Black, Asian, Others)
    • Mothers Marital Status (Married, Single)
    • Mothers Age
    • Mothers education (Mother’s years of education)
    • Maternal Smoking (Mother’s number of cigarettes smoked per day)
    • Maternal Drinking (Mother’s number of alcoholic drinks per week)
    • Adequacy of Prenatal Care: Kotelchuck Index (Four categories, 1-Inadequate, 2-Intermediate, 3-Adequate, 4-Adequate Plus)
Data Sources and Methods

Variables

• Independent Variables:
  – Ambient Air pollution variables
    • Daily average 8-hr maximum O3 concentrations by census tract
    • Daily average PM2.5 concentrations by census tract
  – Three Community-level (census tract) socioeconomic and urbanization variables
    • Income: Family Median Income
    • Education Attainment: Rate of people who have high school and above education
    • Urbanization: Percentage of Urban Land
Data Sources

• **Individual birth and maternal variables**
  – Derived from the birth certificate data (BCD) collected by the state Vital Records Office in Atlanta, Georgia. Only the year 2000 Data were used.

• **Census-Tract Level O3 and PM2.5**
  – USEPA Fused Air Quality Predictions Surface (AQPS)
  – Calculated from the Community Multiscale Air Quality (CMAQ) output and the local and national air quality monitoring data by EPA

• **Census-Tract Level Socioeconomic Factors:**
  – Derived from the US Census 2000

• **Census-Tract Level Percentage of Urban Land**
  – Calculated from the 2001 Georgia land use data, that was originally created by the Natural Resources Spatial Analysis Laboratory at University of Georgia from Landsat Imagery.
The original dataset records 132,286 births. Only all the live, singleton (one fetus carried during a single pregnancy event), mother received prenatal care, births were selected. The records with missing data in any variables were excluded.

Every individual PTB was linked to its individual factors and the air pollution, socio-economic, urbanization variables in the census tract where the maternal residency of the birth is located.
Modeling Techniques

- Both GLR (Global Logistic Regression) and GWLR models are performed.
  - Using PTB as dependent variable, O3 and PM2.5 concentrations as independent variables, adjusted for the nine individual-level variables and the three community SES and urbanization variables.
Model Results Interpretation

• Model results interpretation
  – Comparing the results from GLR and GWLR
  – Mapping the local statistics results (regression coefficient) from GWLR
    • To show how the relationship for each pollutant and PTB changes over space
  – Comparing the community-level socioeconomic and urbanization variables among the cases with different significances in the relationship of an air pollutant with PTB
    • To understand how the relationship is affected by the socioeconomic and urbanization characteristics of communities
Results

Descriptive Statistics on the distribution of term births and PTB in different air pollution levels

<table>
<thead>
<tr>
<th>Variables</th>
<th>All Births (n = 116,112)</th>
<th>Term Births (n = 104,667)</th>
<th>PTB (n = 11,445)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>% of all births</td>
<td>n</td>
</tr>
<tr>
<td><strong>$O_3$ (ppb)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1 (38.33–41.45)</td>
<td>28,948</td>
<td>25,940</td>
<td>89.61</td>
</tr>
<tr>
<td>Q2 (41.45–42.73)</td>
<td>29,061</td>
<td>26,092</td>
<td>89.78</td>
</tr>
<tr>
<td>Q3 (42.73–43.97)</td>
<td>28,989</td>
<td>26,232</td>
<td>90.49</td>
</tr>
<tr>
<td>Q4 (43.97–48.99)</td>
<td>29,114</td>
<td>26,403</td>
<td>90.69</td>
</tr>
<tr>
<td><strong>$PM2.5$ ($\mu g/m^3$)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1 (10.66–13.67)</td>
<td>29,020</td>
<td>25,867</td>
<td>89.14</td>
</tr>
<tr>
<td>Q2 (13.67–16.20)</td>
<td>29,087</td>
<td>26,087</td>
<td>89.69</td>
</tr>
<tr>
<td>Q3 (16.20–16.99)</td>
<td>28,993</td>
<td>26,554</td>
<td>91.59</td>
</tr>
<tr>
<td>Q4 (16.99–18.88)</td>
<td>29,012</td>
<td>26,159</td>
<td>90.17</td>
</tr>
</tbody>
</table>

* The differences in the percentage of PTB among different categories have been tested by Pearson Chi-Square Tests, and $p < 0.001$ for both pollutants.

• Surprisingly, PTB rate is the highest in the lowest quantile for both $O_3$ and PM2.5
Results

Results of global logistic regression on PTB and air pollutants

<table>
<thead>
<tr>
<th>Variables</th>
<th>B</th>
<th>Std. Error</th>
<th>Sig.</th>
<th>Odds Ratio (OR)</th>
<th>OR C.I. (95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>O₃ (ppb)</td>
<td>−0.024</td>
<td>0.007</td>
<td>0.001</td>
<td>0.976</td>
<td>0.962–0.990</td>
</tr>
<tr>
<td>PM₂.₅ (µg/m³)</td>
<td>−0.017</td>
<td>0.007</td>
<td>0.021</td>
<td>0.983</td>
<td>0.970–0.998</td>
</tr>
</tbody>
</table>

Note: Adjusted for individual-level birth and maternal socioeconomic, demographic, behavioral and lifestyle variables and community SES and urbanization variables.

Significant negative relationship found for PTB and both air pollutants.
Spatially Varying Relationships between PTB and Air Pollutants Discovered by GWLR

Descriptive Statistics on GWLR Local Correlation Coefficients and Odds Ratio for Air Pollutants

<table>
<thead>
<tr>
<th>Variables</th>
<th>Correlation coefficients</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min</td>
<td>Median</td>
</tr>
<tr>
<td>$O_3$ (ppb)</td>
<td>-0.5980</td>
<td>0.0408</td>
</tr>
<tr>
<td>PM$_{2.5}$ (µg/m$^3$)</td>
<td>-0.0101</td>
<td>0.0171</td>
</tr>
</tbody>
</table>

Note: Adjusted for individual-level birth and maternal socioeconomic, demographic, behavioral and lifestyle variables and community SES and urbanization variables.

Both positive and negative relationships found for PTB and both air pollutants by GWLR
Spatial Distribution of the Local Correlation Coefficients for O3 and PM2.5 in the GWLR Model
Spatial Distribution of the Significance Level for the Relationship between PTB and Air Pollutants in the GWLR Model
Comparison of Community SES and Urbanization Level among Births with Different Significances in the Associations of PTB with Air Pollutants

<table>
<thead>
<tr>
<th>Significance Type</th>
<th>N (number of births)</th>
<th>% N</th>
<th>Percentile</th>
<th>Family Median Income ($)</th>
<th>High School and Above (%)</th>
<th>Percentage of Urban Land (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>With O₃</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negatively Sig.</td>
<td>27,831</td>
<td>24.0</td>
<td>Mean</td>
<td>51,869</td>
<td>78.0</td>
<td>30.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P25</td>
<td>36,842</td>
<td>68.9</td>
<td>7.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Median</td>
<td>47,401</td>
<td>79.4</td>
<td>18.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P75</td>
<td>62,750</td>
<td>88.4</td>
<td>48.81</td>
</tr>
<tr>
<td>Not Sig.</td>
<td>76,508</td>
<td>65.9</td>
<td>Mean</td>
<td>49,497</td>
<td>77.6</td>
<td>47.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P25</td>
<td>35,455</td>
<td>68.0</td>
<td>12.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Median</td>
<td>44,475</td>
<td>78.6</td>
<td>51.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P75</td>
<td>58,886</td>
<td>88.4</td>
<td>74.92</td>
</tr>
<tr>
<td>Positively Sig.</td>
<td>11,774</td>
<td>10.1</td>
<td>Mean</td>
<td>44,970</td>
<td>74.1</td>
<td>29.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P25</td>
<td>36,017</td>
<td>66.7</td>
<td>9.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Median</td>
<td>44,621</td>
<td>74.7</td>
<td>18.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P75</td>
<td>52,298</td>
<td>82.5</td>
<td>43.50</td>
</tr>
<tr>
<td>With PM₂.₅</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negatively Sig.</td>
<td>18,480</td>
<td>15.9</td>
<td>Mean</td>
<td>59,070</td>
<td>79.3</td>
<td>33.41</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P25</td>
<td>43,146</td>
<td>71.2</td>
<td>11.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Median</td>
<td>54,844</td>
<td>80.6</td>
<td>24.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P75</td>
<td>71,675</td>
<td>90.1</td>
<td>54.05</td>
</tr>
<tr>
<td>Not Sig.</td>
<td>80,169</td>
<td>69.1</td>
<td>Mean</td>
<td>49,262</td>
<td>77.5</td>
<td>46.84</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P25</td>
<td>35,402</td>
<td>68.0</td>
<td>12.84</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Median</td>
<td>44,475</td>
<td>78.5</td>
<td>50.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P75</td>
<td>58,836</td>
<td>88.3</td>
<td>74.77</td>
</tr>
<tr>
<td>Positively Sig.</td>
<td>17,463</td>
<td>15.0</td>
<td>Mean</td>
<td>41,174</td>
<td>74.5</td>
<td>24.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P25</td>
<td>32,755</td>
<td>65.3</td>
<td>6.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Median</td>
<td>38,306</td>
<td>73.0</td>
<td>10.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P75</td>
<td>48,021</td>
<td>83.4</td>
<td>32.76</td>
</tr>
</tbody>
</table>

- O₃ and PM₂.₅ are more significant factors for PTB in low SES rural communities
- Not significant in urban communities
Conclusions

• Major findings:
  • Significant positive, significant negative, non-significant relationships between PTB and both O3 and PM2.5 are all present in Georgia
  • Both air pollutants are important PTB risk factors in only small parts of Georgia, particularly in rural communities with low SES
  • PTB is not significantly related to either air pollutant in urban communities.

• Contributions of this study:
  – Extend the application of GWLR to explore the spatially varying relationships between PTB and air pollution
  – Analyze the effect of community SES and urbanization level on the relationships
  – Help police makers to target the local cause of PTB
  – The methodology can be adopted by health scientists to examine the spatially varying relationships between risk factors and health issues
Thank You

For the 2019 Transportation, Air Quality, and Health Symposium

Feb. 20, 2019

Austin, TX