
Tabitha Julien, MPH
Division of Disease Control
New York City Department of Health and Mental Hygiene
Pneumonia in NYC

- Pneumonia is a mild to severe inflammatory condition of the lungs.

- Pneumonia and influenza (P&I) are the 3rd leading cause of death in NYC but 8th nationwide.

- Numerous studies have focused on individual-level predictors, few have explored the influences of:
  - human behavior
  - housing policy
  - industrial zoning
  - environment
Objective

Understand the association between neighborhood-level socio-environmental predictors and pneumonia-associated hospitalizations (PAH) in NYC
Methods
Data Sources (1)

1) New York Statewide Planning and Research Cooperative System (SPARCS)
   • Key variables:
     • Date of admission, principal and secondary diagnosis codes

2) NYC Department of City Planning Neighborhood Tabulation Area (NTA) Shapefile
   • NTAs were created by the aggregations of census tracts with a minimum population of 15,000
   • 188 residential NTAs in NYC

3) NYC Open Data
   • Social and demographic data
Data Sources (2)

4) Bureau of Environmental Surveillance and Policy
   • Environmental data

5) Bureau of Chronic Disease Prevention and Tobacco Control
   • Tobacco data
Study Population

All NYC residents during 2010–2014

Outcome:
- Inpatient pneumonia-associated hospitalizations rates at the NTA
  - Rates calculated using the 2010 census population estimates for NYC
Predictors

Selected based on data availability at the NTA level...

- **Environmental (N=8):** industrial areas, age of housing, particulate matter, black carbon, nitric oxide, nitrogen dioxide, ozone, and sulfur dioxide.

- **Social (N=13):** education, median household income, liquor stores, tobacco retailers, healthcare facility, etc.

- **Demographic (N=24):** race, gender, age, etc.
Accounting for spatial variability in associations

Assumption:

• Strength of association is constant across NYC
• Predictors will equally impact health outcomes

• What if place impacts how these predictors influence health outcomes?

• What if industrial areas in Brooklyn and Queens emit more hazardous pollutants than Staten Island?
Step 1: City-wide model

City-wide model (linear regression) assumes that associations (coefficients) and goodness of fit ($R^2$) are constant over space

- Model predictors were identified based on previous literature and significance

Model Statistics

- $R^2$ value is an index from 0–1
  - 1 indicates a perfect fit
- Coefficient describes the strength of association between predictor and outcome
- Statistical software: R
Step 2: Geographically weighted regression model

Geographically weighted regression (GWR) allows these associations to vary over space

• Coefficients and $R^2$ values are calculated for each neighborhood and each predictor

• GWR allows us to fit a linear regression model for each neighborhood for each predictor

• Geographic Unit: NTA

• Statistical software: ESRI ArcGIS, Spatial Statistics Tools
Step 3: K-means clustering

- K-means clustering identifies groups based on similarity
- Allows us to identify neighborhoods with similar coefficients
- Geographically summarizes most important predictors
- Statistical software: ESRI ArcGIS, Spatial Statistics Tools
Results
Average annual pneumonia-associated hospitalization rates (per 100,000 population) NYC, 2010–2014
### Linear Model Summary Statistics

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Description</th>
<th>Coefficient</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industrial (%)</td>
<td>Industrial areas</td>
<td>12.36</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

- Strength of association: 12.36

- 1% increase in industrial area per NTA = on average, 12.36 increase in PAH per 100,000 population
### Linear Model Summary Statistics

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Description</th>
<th>Coefficient</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industrial</td>
<td>Industrial areas</td>
<td>12.36</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Race (%)</td>
<td>White persons</td>
<td>-1.91</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

- Strength of association: -1.91
- 1% increase in white persons per NTA = on average, 1.91 decrease in PAH per 100,000 population

*1 hectare = 10,000 sq. m ~ 1 football field*
## Linear Model Summary Statistics

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Description</th>
<th>Coefficient</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industrial (%)</td>
<td>Industrial areas</td>
<td>12.36</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Race (%)</td>
<td>White persons</td>
<td>-1.91</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Age (%)</td>
<td>Persons &lt;5 and ≥65 years</td>
<td>19.67</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Education (%)</td>
<td>Persons earning less than a high school diploma</td>
<td>2.66</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Healthcare Facility (per 100 hectares(^{†}))</td>
<td>Density of healthcare facilities</td>
<td>4.35 X 10^3</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

\(^{†}\) 1 hectare = 10,000 sq. m ~ 1 football field
Linear regression helped us identify citywide associations but...

GWR can help us find how the strength of these associations vary across neighborhoods!
Persons age <5 and ≥65: positively associated with PAH rates

Coefficients
- 26.50–52.94
- 18.70–26.49
- 10.98–18.69
- -6.40–10.97
- Non-residential
Density of healthcare facilities: positively associated with PAH rates

Coefficients
- Dark purple: 10.87–94.21
- Purple: 6.56–10.86
- Light purple: 4.44–6.55
- Lightest purple: -10.23–4.43
- White: Non-residential
Earning less than a high school diploma: positively associated with PAH rates

Coefficients
- Dark purple: 2.28–9.19
- Magenta: 0.55–2.27
- Medium purple: -0.57–0.54
- Light purple: -4.08 – -0.58
- White: Non-residential
White: negatively associated with PAH rates

Coefficients

-11.09–-3.71
-3.70–-2.47
-2.46–-1.70
-1.69–1.31
Non-residential
Industrial area: positively associated with PAH rates

Coefficients
- Dark purple: 22.18–70.05
- Purple: 10.58–22.17
- Light purple: 3.09–10.57
- Pale purple: -22.44–3.08
- White: Non-residential
GWR further explains associations

- $R^2$ value is an index from 0–1
- Value that indicates goodness of fit
  - How well the predictors in the model explains pneumonia-associated hospitalization rates

<table>
<thead>
<tr>
<th>Fitness Parameters</th>
<th>Linear</th>
<th>GWR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjusted $R^2$</td>
<td>0.49</td>
<td>0.68</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.51</td>
<td>0.79</td>
</tr>
</tbody>
</table>
Local $R^2$ values obtained from GWR model of PAH rates in NYC, 2010–2014

- Dark red: $0.67–0.98$
- Orange: $0.59–0.66$
- Light orange: $0.44–0.58$
- Yellow: $0.18–0.43$
- White: Non-residential
To help synthesize GWR results—we can use K-mean clustering to summarize neighborhood-level associations...
K-means Clustering: PAH associations vary by neighborhood
What is potentially driving PAH at the neighborhood-level?

• GWR is an exploratory analysis

• We suggest further exploration of...
  • Less educated population in southern Staten Island
  • Healthcare facility density in southern Staten Island and NYC overall
  • Racial disparities in Brooklyn, Bronx, and Staten Island
  • Neighborhoods that have industrial areas in Queens and Brooklyn
Limitations

- We cannot predict individual level risk for pneumonia-associated hospitalizations

Biases
- Accuracy of discharge records
  - Under and over reporting of pneumonia-associated hospitalizations
  - Diagnosis codes are assigned after the patient is discharged from hospital
- However, diagnosis codes were chosen based on a previous study to overcome coding biases

- Outpatient hospitalizations were not captured in this study
Current pneumonia studies

• Descriptive analysis of pneumonia deaths in NYC

• Comparing pneumonia-associated hospitalization rates in NYC to rates across the United States

• Investigation of pneumonia by setting of acquisition
Next Steps? (1)

- Comparison of pneumonia-associated hospitalizations that result in a fatal outcome in NYC versus those that do not

- Chart review of all pneumonia deaths at New York Presbyterian/Columbia University Medical Center over a recent time period

- Comparison of pneumonia deaths in NYC vs the United States
Next Steps? (2)

- In-depth review of sepsis and its relationship to pneumonia deaths

- Discuss findings with subject matter experts in pneumonia across NYC medical schools and at the Centers for Disease Control and Prevention (CDC)
Thank You!

Data Unit
- David E. Lucero
- Pallavi Kache
- Christopher Gu

Informatics Unit
- Rachel Corrado
- Angel Aponte

Disease Control Administration
- Neil Vora
- John Rojas
- Jay K. Varma

GIS Services
- Susan Resnick

Bureau of Epi Services
- Gretchen Culp
- Candace Mulready-Ward

Bureau of Environmental Surveillance and Policy
- Sarah Johnson

Bureau of Alcohol Drug Treatment and Prevention
- Shannon Farley

Bureau of Public Health Engineering
- Rodolfo Perez

NYC Department of City Planning
Questions?

Contact Information:

Tabitha Julien, MPH
Division of Disease Control
tjulien1@health.nyc.gov