

Harmonizing Multi-Level Public Health Data to Guide Lung Cancer Screening in New York City

M. Lipsman, K. Burke, P. Tehranifar

Herbert Irving Comprehensive Cancer Center, Columbia University Irving Medical Center

1. Background

Lung and bronchus cancer remain leading causes of cancer-related death in New York City (NYC), with smoking as a primary risk factor. Access to lung cancer screening is uneven across communities, with only about one in five eligible individuals receiving lung cancer screening in New York state.

Identifying neighborhoods with the highest burden of smoking and lung cancer outcomes can inform the deployment of targeted screening interventions. This project aimed to create a census tract-level risk scoring system to guide resource allocation and outreach, and to identify highest priority census tracts.

2. Goals

The goal was to integrate multiple datasets to create a census tract-level composite risk score for lung and bronchus cancer in NYC. Objectives were to:

- Collect and harmonize data on adult current smoking prevalence, lung/bronchus cancer incidence, and lung/bronchus cancer mortality
- Translate disparate geographic units (census tract, aggregated census tract, community district) into a consistent census tract-level dataset
- Create a scoring system to identify areas of highest need for lung cancer screening outreach

Target outcomes included a ranked list of 2,227 NYC census tracts and highlighting top tracts for potential mobile van deployment.

3. Solutions and Methods

We used three datasets: smoking prevalence by census tract (Behavioral Risk Factor Surveillance System, 2022), cancer incidence by aggregated census tract (New York Cancer Registry, 2017–2021), and cancer mortality by community district (NYC Department of Health EpiQuery, 2021). Incidence and mortality data were converted into census tract-level estimates for consistency. Each variable was assigned a score from 1 to 4, with higher values corresponding to higher burden. Cut points were defined by dividing the range of values into quartiles. For each tract, the three variable scores were summed to generate a composite “risk score” (range 3 to 12). Census tracts were ranked based on this score, providing a data-driven prioritization scheme for screening resources.

When converting lung/bronchus cancer incidence data from aggregated census tracts to individual census tracts, some data were missing, so we used the average of all known incidence rates in the corresponding community district.

4. Outcomes

The analysis produced a scoring system covering all 2,227 NYC census tracts. Tracts with the highest composite scores were identified as high-priority areas for mobile lung cancer screening. Twenty-nine census tracts had the highest score of 10, representing neighborhoods with high smoking prevalence, cancer incidence, and mortality. These census tracts were in Harlem and Midtown/Flatiron/Union Square (Manhattan), Ocean Hill/Brownsville (Brooklyn), South Shore and Mid-Island (Staten Island), and the Rockaways (Queens). Results are now informing lung cancer screening

van placement and education outreach in six neighborhoods. The mobile van provides low dose computed tomography scans directly in communities, reducing travel barriers and competing priorities.

5. Lessons Learned and Future Directions

This scoring system demonstrates the feasibility of integrating heterogeneous datasets to inform cancer education and screening outreach. Key challenges included reconciling data reported at different geographic scales and ensuring comparability across metrics. Future work will refine the methodology by incorporating additional risk factors like socioeconomic status and exploring dynamic updates as new surveillance data becomes available. The approach may also be adapted for mobile health interventions in NYC and beyond.

Figure: Mapping of Total Risk Scores by NYC Census Tract

