**Integration of Low-Cost Pollution Sensor Data for Neighborhood-Level Health Studies in Philadelphia**

Colin Christie, Sherrie Xie, B.S., Avantika Diwadkar, M.S., Rebecca E. Greenblatt, B.S., Blanca E. Himes, Ph.D.

Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

**Introduction.** Monitoring particulate matter (PM) is important because exposure has been associated with various negative health outcomes. While PM levels have been measured for decades at discrete regulatory stations, concern for PM’s effects on health along with broad demand for accessible environmental monitoring have led researchers and manufacturers to develop low-cost, portable pollution sensors in recent years. These sensors offer an alternative method for PM monitoring with the advantage of taking measurements anywhere and at any time, despite having less accuracy than regulatory monitors. In Philadelphia, PM sensors have been utilized by citizen scientists, researchers, and non-profit groups to assess neighborhood-level exposures to air pollution. Other health-related variables, including neighborhood socioeconomic status, crime incidence, and exposure to vehicular traffic, exhibit high geospatial variability in Philadelphia and have been linked to negative health outcomes such as asthma exacerbations. Understanding the drivers of geospatially varying health disparities in Philadelphia thus requires fine-scale, neighborhood-level data on the physical and social environments. We created Sensor-based Analysis of Pollution in the Philadelphia Region with Information on Neighborhoods and the Environment (SAPPHIRINE), an app that integrates sensor-based PM data from online databases, PM measurements taken by our lab, and other geospatial data on Philadelphia neighborhoods into an interactive, highly specific visualization tool to facilitate neighborhood-level health studies.

**Methods.** SAPPHIRINE was created using the R shiny library. Code used to create the app, as well as links to the data sources used, are available at [https://github.com/HimesGroup/sapphirine](https://github.com/HimesGroup/sapphirine). The app can be accessed at [http://sapphirine.org](http://sapphirine.org). The app includes measurement data for PM$_{2.5}$, PM$_{1.0}$, and PM$_{10}$ concentrations as well as temperature, humidity, timestamp, and location. Some of these measures were taken by our lab with AirBeam sensors in Philadelphia between October 23, 2017, and July 26, 2018, and the rest were downloaded from crowdsourced databases, spanning a timeframe from June 2015 to May 2019. Daily summary data for PM$_{2.5}$ concentration from the EPA were downloaded and, using inverse-square-distance-weighted interpolation, rendered as a color-scaled raster layer projected over a map of Philadelphia. Tabular datasets of publicly available Philadelphia crime incidents, Area Deprivation Index (ADI), and Annual Average Daily Traffic (AADT) were also downloaded and incorporated into SAPPHIRINE.

**Results.** SAPPHIRINE allows users to generate custom maps with color-scaled raster layers of bins that display PM$_{1.0}$, PM$_{2.5}$, PM$_{10}$, temperature, humidity, crime, poverty, and traffic variables. Adjustable parameters include variable type, grid resolution, location, timestamp, and sensor ID. For sensor-based data types, all available data points falling within each bin are averaged, whereas crime incidents are summed. For ADI and AADT, pre-existing raster layers are imported and resampled using bilinear interpolation. Measurement densities for sensor-based data types are plotted as log-transformed sums. Comparison of sensor-based versus EPA PM$_{2.5}$ measures found that the former covered a larger geographical area and captured finer-scale fluctuations with a greater range of values. For example, while there was little apparent correlation between traffic volume and PM$_{2.5}$ concentrations when PM$_{2.5}$ was estimated from all available sensor data, such a relationship was observed when PM$_{2.5}$ was estimated from only measurements taken between 4pm and 6pm.

**Conclusion.** We created SAPPHIRINE, an online web-application that integrates sensor-based pollution measures alongside other geographically informative data in the Greater Philadelphia Area to facilitate its use for research, policy, and citizen-science efforts. Data visualized in SAPPHIRINE can be customized according to several user-defined parameters, enabling tailored queries for a variety of purposes, including the linking of environmental and social data to health outcomes.