# ACHD: Mon Valley Air Toxics Study

Sage Betko, Michelle Koo, Minseo Kwon (Advisor: Peter Freeman)

## Introduction

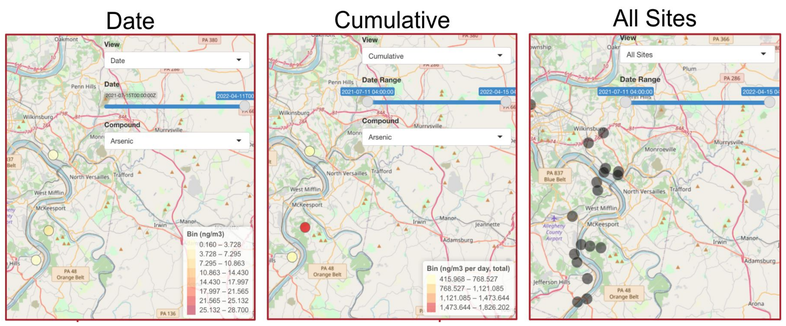
Outdoor air pollution is a significant contributor to the burden of disease worldwide, and is causally linked to cancer incidence and mortality. The Allegheny County Health Department (ACHD), as part of the ongoing Mon-Valley Air Toxics Study (2020-present), has collected data of air pollution by installing dozens of monitoring stations throughout the Monongahela Valley southeast of Pittsburgh. The Mon Valley is an industrial area with known air quality issues and is recognized by the Environmental Protection Agency (EPA) as an Environmental Justice area, as its residents are burdened by environmental hazards such as poor air quality. This study aims to develop a browser-accessible interactive map that allows one to visualize pollutants and to estimate cancer risks in the region.

## Data Description

- **Data sources:** Volatile organic compounds (VOCs) including benzene, toluene, ethyl-benzene, and 3 xylene isomers are collected from 16 stations situated throughout the Mon Valley at a 1-in-2 week frequency, in addition to a 1-in-3 day measurement station in Liberty. A variety of other pollutants including particulate matter (PM2.5 and PM10), sulfur dioxide, and carbon dioxide are collected from eight hourly monitoring sites. Weekly concentrations of metals including arsenic, beryllium, cadmium, chromium, lead, manganese, nickel, and cobalt are collected at sites in Clairton, Liberty, and North Braddock [1].
- **Unified dataset:** 604,169 standardized records with concentration, unit of measurement, time of collection, latitude and longitude, compound or compound species name, and a reference to the originating file. In addition to the eight primary fields associated with every record, other information specific to certain measurement sites or modes of measurement are included in the unified dataset for completeness. These include wind speed, wind direction, sensor-specific minimum detection limits, and miscellaneous quality-assurance flags.

## Geographic Visualization Web Application

- We used the Leaflet shiny R package in order to create the interactive and responsive application that allows one to visualize pollutants and to estimate cancer risks in the region – including three different views: `Date`, `Cumulative`, and `All Sites`.
- The `Date` and `Cumulative` view shows the geographic summary of a specific compound and the `All Sites` shows an overview of all compound measurements.




## Methods

I. **Computing Cumulative Cancer Risk**. We observe the additive effects of chronic inhalation cancer risk by using mean pollution measurements and the EPA's compound-specific unit risk risk estimate (URE) for lifetime chronic exposure [2].

**Cancer Unit Risk Estimate (URE)**
**× (Average Exposure)**
**= 1 in 1,000,000 cancer risk from pollutant**

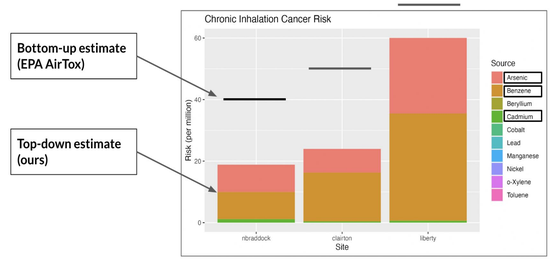
Compound-Specific Cancer Risk

II. **Interpolating in Space and Time**. Not every air monitoring station every compound; since cumulative cancer risk is a sum of compound specific risks, we need a method for estimating concentrations at arbitrary points and times to get a complete risk profile of the region. We perform linear interpolation to standardize sampling rates across sites before applying a Kriging algorithm in order to interpolate and predict values over a dense spatial grid with 50x50 meter resolution.

## Results

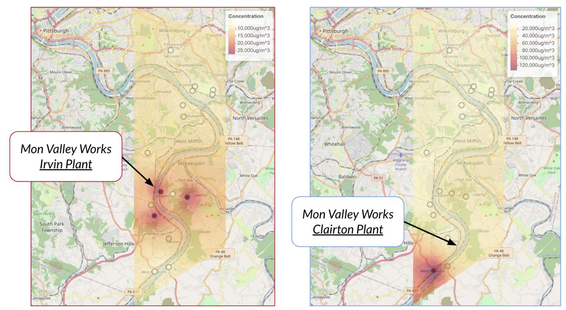
### I. Result of Cancer Risk Assessment

- Liberty has more than two times higher cumulative cancer risk compared to Clairton and North Braddock.
- We found that our bottom-up cancer risk estimates (derived from *in situ* measurements) are consistently lower than the EPA's top-down estimates (based on atmospheric modeling of industry-reported emission rates) from the 2018 AirTox Assessment, but are in similar relative proportions across key sites [3].




### II. Result of Interpolation

- Pollutant concentrations, especially of VOCs, are highly localized within the study region and are clustered near industrial sites.
- Benzene particularly exhibits high spatial variability. The three sites with elevated mean concentrations of >4 μg/m$^3$, in Liberty, Glassport, and Clairton, are each separated by adjacent sites averaging <2 μg/m$^3$.




Cumulative daily **benzene** | Cumulative daily **toluene**

## Discussion/Conclusions

- Toxin measurements are above zero (on average), indicating further public health analysis should be done.
- Analyses indicate arsenic and benzene are the driving factor for regionally elevated chronic cancer risk.
- The high degree of spatial variability is localized to specific communities in the Mon Valley, and suggests that the correlation structure of the spatial interpolation may be strongly dependent on local topography or weather conditions (e.g., altitude or micro-climate variation).
- This analysis substantiates the necessity for the ACHD's "top-down" approach owing to deviations from prior national studies based on bottom-up estimates.

## Future Work

- Implement a unified data collection method for easy integration into Leaflet app and to keep app current with new data.
- Integrate spatial interpolation results on Leaflet Shiny application and incorporate meteorological data such as wind speed and direction, as well as geographical priors into the Kriging algorithm.
- Summarize target-organ specific disease hazards (non-cancer assessment).

## References

1. Air Quality Reports and Studies | Allegheny County. (2022, December 12). Retrieved from `https://www.alleghenycounty.us/Health-Department/Resources/Data-and-Reporting/Air-Quality-Reports/Air-Quality-Reports-and-Studies.aspx`
2. US Environmental Protection Agency. (2022). Risk Assessment for Carcinogenic Effects. US EPA. Retrieved from `https://www.epa.gov/fera/risk-assessment-carcinogenic-effects`
3. US Environmental Protection Agency. (2022). Air Toxics Screening Assessment. US EPA. Retrieved from `https://www.epa.gov/AirToxScreen`