Samrie Beshah, PGY-2
Faculty Mentor: Stephanie Lovinsky-Desir, MD, MS
July 22, 2022

Title: Pollution and Lung health in Active Youth: PLAY 1

Background

The term social determinants of health describes the principle that where someone lives impacts health. Social determinants of health can be studied through examining built and social environments. Built environments are defined as the buildings, transportation systems, access to water, as well as recreational spaces that make up someone's environment[1]. Social environments are defined as the stability of social connections, demographics, and social capital that impact one's environment[2]. By focusing on the built and social environments, geographical and survey data can be used to demonstrate the relationship between where patients live and pediatric health outcomes.

Geographic information systems (GIS) and the US Census data provide a variety of variables (built and social environments) about the communities in which children live. The research team has used these technologies to identify built and social environment characteristics for participants of the Urban Environment and Childhood Asthma (URECA) longitudinal birth cohort. These data were collected using two distinct methodologies: a centralized approach in the Pollution and Lung health in Active Youth (PLAY 1) study and a decentralized approach under the School vs. Home Environmental Exposures and Respiratory Outcomes (SHEERO) study. The centralized data from the PLAY 1 study has limitations due to the need to transfer protected health information (PHI) to a centralized site, which differs from the decentralized data from the SHEERO study which does not require transfer of PHI as shown in the diagram below. While the centralized approach has been an effective tool for obtaining data, there are privacy limitations to centralized data collection including using in-house computers, secure servers, ways of analyzing data and concern for HIPAA violations. With the PHI from the decentralized data removed, there can be more easily shared data in multi-site studies as well as use of different online geocoding site (i.e. Google Maps), which cannot be used with centralized data[3]. These concerns can be limited by using a decentralized data set that takes the geographic data and links it to the US Census and other survey data without attaching that information to specific patients or locations. By comparing the outcomes of the decentralized data set and the centralized data set, there is an opportunity to further learn about the cohort in a safer way with less limitations.

---

[1] Basic Information about the Built Environment | US EPA." *U.S. Environmental Protection Agency*, U.S. Environmental Protection Agency, 13 Apr. 2017, https://www.epa.gov/smm/basic-information-about-built-environment.

[2] Institute of Medicine and National Research Council. 2013. *U.S. Health in International Perspective: Shorter Lives, Poorer Health.* Washington, DC: The National Academies Press. https://doi.org/10.17226/13497.

[3] Brokamp C, Wolfe C, Lingren T, Harley J, Ryan P. Decentralized and reproducible geocoding and characterization of community and environmental exposures for multisite studies. *J Am Med Inform Assoc.* 2018;25(3):309-314. doi:10.1093/jamia/ocx128
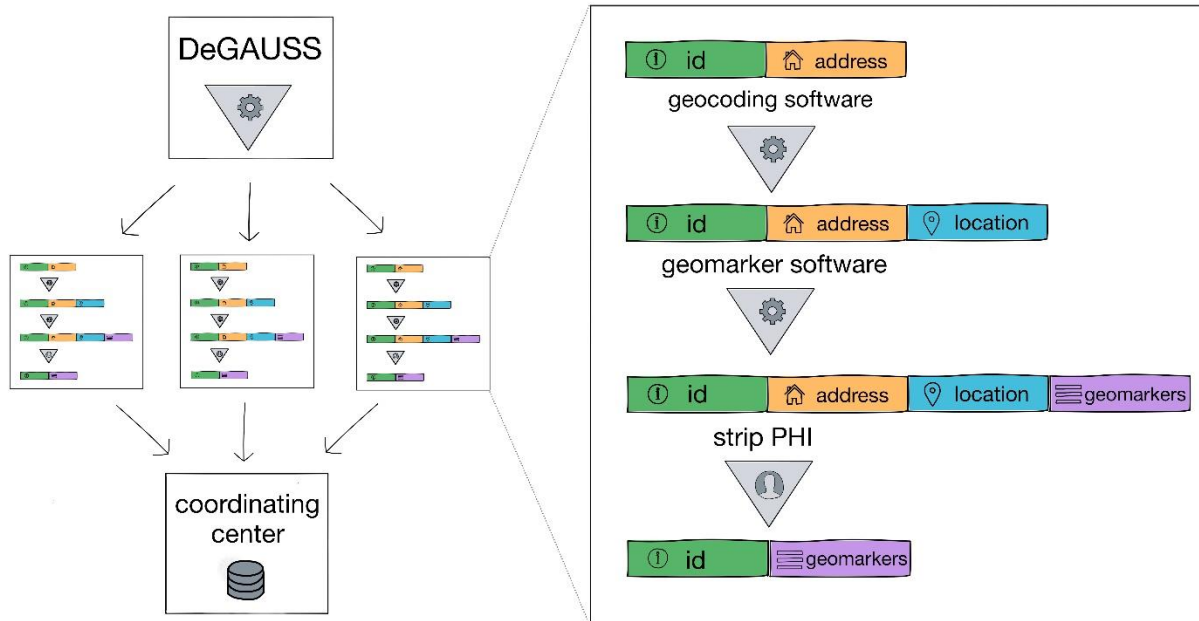
Diagram 1: Visual representation of the decentralizing method of geocoding that allows members of the cohort to be attached to the geomarkers without the use of PHI.

Aims:

Understand the relationship between built and social environments and the impact on asthma.

Methods:

This is an observational study that will utilize data that have been collected for a larger study. We will compare of two methodologies of data analysis of built and social environments to see if health outcomes between the centralized and decentralized data provide the same results.

We will first compare variables of neighborhood built and social environments that were collected using the centralized vs. the decentralized approaches using paired T-tests for continuous variables and chi-square tests for categorical variables. I will spend time familiarizing myself with the different metrics for assessing neighborhood environment such as which are the best variables to determine socioeconomic status (e.g. household income level vs. percent of population living below federal poverty level or percent of population with high school education vs. percent of population with graduate degree or higher). I will also familiarize myself with the many different comprehensive indexes of built environment exposure including the area deprivation index (ADI) and the childhood opportunity index (COI). I will then work with a statistician to determine which of these variables best predict asthma outcomes in the sample of children represented in our cohort.

[4] Brokamp C, Wolfe C, Lingren T, Harley J, Ryan P. Decentralized and reproducible geocoding and characterization of community and environmental exposures for multisite studies. *J Am Med Inform Assoc.* 2018;25(3):309-314. doi:10.1093/jamia/ocx128