

Urban Health Collaborative

Research and Data Core

Brief description of available data

Introduction

The primary objective of the Urban Health Collaborative (UHC) Research and Data Core at the Dornsife School of Public Health (SPH) is to support work that characterizes health in cities and to conduct research on health determinants and on the impact of policies. As part of its activities the Data and Research Core of the UHC has begun to compile data on health determinants and health outcomes for the Philadelphia region.

This document briefly outlines data that the UHC has compiled to date. The purpose of this document is to orient potential Drexel users to the available data so that they can assess whether these data could be useful for student projects, pilot analyses/grant writing, or in-depth studies. [Please contact the UHC](#) if you: 1. want more details on available data, 2. want to request new/additional data for your research or evaluation studies, 3. have data that you'd like to add to the data available. The UHC has written additional documents that describe the data in more detail; the detailed documents will be made available to the researcher as needed. The UHC continues to work to expand and update data available so potential users should contact UHS staff for more information if a domain they are interested in is not currently listed. The UHC is also happy to receive suggestions on additional domains or data sources that could be incorporated.

A. Data geographic extent and level of aggregation

The primary geographic focus of the UHC data compilation effort to date is the Philadelphia metropolitan area. The UHC defines this as the Philadelphia-Camden-Vineland Combined Statistical Area (PCV-CSA) and Mercer County, NJ (where Trenton, NJ is located) to make up the Philadelphia-Camden-Vineland-Mercer area (PCVM) and includes the following counties: Berks, Montgomery, Bucks, Burlington, Mercer, Camden, Gloucester, Cumberland, Salem, New Castle, Cecil, Chester, Delaware, and Philadelphia. Some data are available only for the Philadelphia-Camden-Wilmington metropolitan area (PCW) which includes all counties listed above except for Berks, Mercer, and Cumberland.

Most of the data listed below are available for the PCVM, include more than one year of data, and have been spatially indexed to census tracts (i.e., attribute densities within a tract or percent of tracts covered by a feature or tract centroid). Alternate spatial representations can potentially be derived upon request (features within ½ mile, 1 mile, 3 mile, etc.). In order to simplify this document, details on spatial and temporal extent are not provided (with a few exceptions).

B. Population characteristics and area-level exposures

US Census small area data have been compiled for decennial Censuses 1980, 1990, 2000, 2010 and the American Community Survey 5-year moving averages representing years 2005-09, 2007-11, 2008-12, and 2010-14. Census data are the basis for many domains below.

Available data at the UHC include the following:

1. Demographics, residential mobility, and 'economic environment'

- **Age, race/ethnic composition and foreign birth.**
- **Residential density and mobility**, such as population density, percent of persons living in same house in the past year and past 5 years.
- **Measures of the "economic environment"**
 - Socio-economic indicators such as educational attainment, unemployment, poverty, income and assets (including vehicle ownership and owner occupied housing), etc.

- Factor analysis based summary measures of socioeconomic status
- Costs of housing and transportation as well as a "location affordability index" (cost of housing and transportation relative to the regional median income).
- Foreclosure data (tract-level monthly counts of foreclosures, per 1000 housing units).

2. Measures that relate to the 'social environment'

- **Perception of neighborhood "social capital"** (people in neighborhood willing to help their neighbors, feeling of belonging in the neighborhood, etc. compiled from the Southeastern Pennsylvania Household Health Survey (SEPAHHS), conducted every other year 2002-2014/15).
- **Perception of neighborhood "safety"** (did not go someplace during the day because you felt you would not be safe, etc, compiled from SEPAHHS).
- **Crime incidents** per 1000 population in Philadelphia by crime type including total all crimes, burglaries, aggravated assaults and incivilities (narcotic/drug law violations, weapons violations, etc.).
- **Racial segregation indices** (computed from census data) that quantify spatial clustering by race/ethnicity and diversity of different race/ethnicities within an area.

3. Measures that relate to the 'built environment' that could impact physical activity (mostly related to active transportation), injuries, and possibly mental health:

- **Population density** (from census years noted above)
- **Land-use mix** which indicates how land is being used (residential, commercial and retail purposes).
- **Public transportation stops**, calculated separately for commuter rail, rapid transit, light rail and buses (buses are only for PCVM).
- **Pedestrian street network** (compiled from existing road files) and tract-level measures representing street intersection density and street network indices (network area and ratio for 800 m buffers around tract centroids).
- **Road traffic**, including annual average daily traffic for tracts and select buffer areas around tract centroids.
- **Pedestrian/bike fatality in motor vehicle crashes:** fatal crash data for PCVM and non-fatal crash data for Philadelphia County.
- **Walk Score, Bike Score, and Transit Score.** These scores for "walkability," "bikeability," and access to public transit, respectively, were created by Redfin Corporation and purchased for select areas within the target region. In particular, Walk Score data have been obtained for a set of about 1000 points distributed across the Philadelphia region to determine the variability across the city and selected surrounding counties. "Transit Score" and "Bike Score" data are more limited and are restricted to Philadelphia County. (Other related data that have been compiled for some areas: "ZIP code Walk Score" and "Street Smart Walk Score".)
- **Park-based recreational facilities** (i.e., basketball courts and playing fields), have been compiled for Philadelphia county.
- **Tobacco retailers:** Location of retailers licensed to sell tobacco in the city of Philadelphia

Other measures not yet created such as distance to- or density of *vacant lots* can be derived from available data (from parcel data) and motor vehicle crashes that do not involve pedestrian/bike.

4. Measures that relate to the 'food environment'. Food environment can impact diet and possibly mental health outcomes. To date, the following data are available:

- **Locations of food stores** that accept Supplemental Nutrition Assistance Program (SNAP) and Electronic Benefits Transfer (EBT).
- **Perception of the food environment** (easy or difficult is it for you to find fruits and vegetables in your neighborhood, etc., compiled from SEPAHHS).

5. Natural (land) environment. Natural features in the environment may affect health through mechanisms that may be independent of physical activity, for example by reducing stress and improving mental health. To date, the following data are available:

- Percent of **open space** (land that is not intensively developed for residential, commercial, industrial or institutional use; 2010).
- **Impervious surface** area within a tract (sum of roads, parking lots, sidewalks, rooftops and other impermeable surfaces), percent of tree canopy within a tract (leaves, branches, and stems of trees that cover the ground when viewed from above).

Other measures not yet created such as *vegetation index* can be derived from available satellite data.

6. Classic environmental exposures: ambient air pollution, industrial facilities, water, soil. Data listed below have been downloaded and measures can be compiled for specific research projects.

- Daily **ambient air pollution**: particulate matter <2.5 micrometers in diameter have been downloaded but additional pollutants are also available.
- Total **emissions** by pollutant from the National Emissions Inventory database.
- National Scale **Air Toxics Assessment modeled estimates**.
- **Road traffic** (annual average daily traffic for tracts and select buffer areas around tract centroids)
- **Location of industrial facilities**

Other measures not yet created that relate to *healthy housing* could be compiled based on licensing data: permit, license, inspection, case and violation data. These could be used in addition to census data available for year housing was built (median year structure built).

C. Health outcome data

1. The Southeast Pennsylvania Household Health Survey (SEPAHHS, Public Health Management Corp). Administered in five counties every other year (Bucks, Chester, Delaware, Montgomery, and Philadelphia). Through an agreement with PHMC, Dornsife SPH currently has access to data for even-years 2002-2014/2015. The UHC modified the data files to create a longitudinal dataset, standardized measures across years as much as possible, created common re-coding or summary variables for use in longitudinal analyses, and derived weighted tract-level means. Geocoding is available to the census tract level, ZIP code, health service areas, health districts, planning districts and other neighborhood definitions as needed.

The survey collects data on:

- **Demographics and SES:** sex, age, household composition, employment status, education, marital status, home ownership, race/ethnicity, immigration status, religion, language, welfare status, income, internet usage, sexual identity, insurance.
- **Self-reported anthropometry, health status:** allergies, arthritis, asthma, cancer, heart condition, diabetes, substance abuse, high blood pressure, high cholesterol, stroke, HIV testing/status, mental health and treatment status, depression scale (60+ only), stress rating, disability, activities of daily living (60+ only), and select medications.
- **Health-related behaviors:** high risk for HIV behaviors, smoking status, smoking cessation attempts, others smoke in car/home, vaping, alcohol use, exercise, sleep rating, use public facilities, fruit/vegetable consumption, soda consumption, sodium use, fast food, TV/computer time, food availability, gambling behavior, pet ownership (60+ only).
- **Health care experience/access:** usual source of care, language barriers, time since last medical visit, visits to ER, rating of health care services, transportation barriers, time since

dental visit, time since eye exam, time since health tests (blood pressure, cholesterol, Pap smear, breast exam, mammogram, osteoporosis, PSA test, colonoscopy), flu shot.

- **Social environment was collected for some years:** guns in home, safe in neighborhood, physical violence, neighborhood participation/involvement, social capital, talk to/see friends and relatives.

2. Vital Records. Birth and death certificate records for Pennsylvania. As of January 2016, the UHC is in the process of obtaining birth and death certificate records 1960-2014 for the entire state of Pennsylvania. The UHC will geocode the addresses provided with the dataset. The UHC will calculate the following outcomes overall and by sex, race/ethnicity, and age group: infant mortality, death rates overall, death by various causes, and life expectancy.

- **Data available on the death records include:** cause of death, underlying cause of death, age, sex, race/ethnicity, marital status, US military veteran status, date and time of death, work place injury, place of birth, place of injury, place of death, usual occupation, type of facility, education, tobacco use contribution to death, pregnancy status, transport accident status.
- **Data available on the birth records include:** adoption status at birth, sex, birth date and time, mother and father age, mother and father education, mother's marital status, mother's live births now living and deceased, date of last live birth, past pregnancy terminations, month began prenatal visits, number of prenatal visits, length of pregnancy, birth weight, plurality, birth order, Apgar score, concurrent illness/conditions, complication of labor/delivery, birth injuries, congenital anomalies, mother and father race/ethnicity, medical risk factors, mother tobacco use, mother alcohol use, obstetric procedures, method of delivery, mother's height, mother's weight pre-pregnancy and at delivery, breastfeeding status.

3. Inpatient and outpatient/ambulatory care discharge data for Pennsylvania. These data are reported to/collected by Pennsylvania Health Care Cost Containment Council (PHC4). Hospitals and freestanding ambulatory surgery centers licensed with the Commonwealth of PA are required to submit data to PHC4. UHC has obtained a pilot dataset for the city of Philadelphia for year 2014 for both the standard inpatient and outpatient/ambulatory datasets. ZIP code of residence will be provided with the data (not address).

- **Inpatient Discharge data** include: admitting diagnosis, discharge status, diagnosis codes from hospital stay (including secondary diagnoses), procedure codes (all procedures), primary payer, all charges/costs, patient age race gender zip code.
- **Outpatient/Ambulatory** include procedure, principal and secondary diagnosis codes, principal procedure and other/secondary procedure codes, primary payer, charges/costs, patient age race gender zip code. Outpatient data doesn't contain all procedures, only those on the mandatory reporting list (see here: <http://www.phc4.org/services/datarequests/procedures.htm>).

The UHC is in the process of identifying other health data. Suggestions are welcome! It is also possible to link any available data in section B to geocoded health data that researchers may already have available. [Please contact the UHC](#) with any questions.