Big Cities Health Inventory (BCHI) data platform TECHNICAL DOCUMENTATION v2

ABOUT THE DATA & METRIC SELECTION

As of April 2024, the Big Cities Health Inventory (BCHI) data platform has 124 metrics. The metrics encompass broad categories of public health importance: Access to Health Services, Chronic Health Conditions, Demographics, Infectious Diseases, Life Expectancy and Deaths, Maternal and Child Health, Mental Health and Substance Use, Climate and Built Environment, Poisoning, Social and Economic Factors, Violence and Injury. These categories were chosen because of their relationship to leading causes of morbidity and mortality in the United States and their role in creating healthier, safer communities.

Platform metrics were selected if they met 'substantive' and 'coverage & standardization' criteria.

> Metrics were selected if they met at least 1 of the following 'substantive criteria': 1. public health relevance, in particular their alignment with CDC's Healthy People goals; 2. ability to serve as benchmarks for new policy initiatives; or 3. can be used to highlight demographic and socioeconomic disparities in health and healthy environments. Note that about one-half of the metrics were in BCHC's prior version of the platform (before this new/enhanced platform was developed).
>
> Metrics were selected if they met all of the following 'coverage & standardization criteria': 1. there is a uniform source of the data, thereby ensuring comparability of the metric across jurisdictions; 2. data are available for at least 50% of the BCHC cities; and 3. data are available for recent years.

Additionally, we prioritized publicly available data (to facilitate ease of updating data and to enable data sharing) and city-level data (as opposed to county-level).

DATA SOURCES

Most of the data in the Big Cities Health Inventory come from publicly available data sources for which city data were available. A few of the sources required data use agreements to release data at the city level. Details about the data sources are given in **Table 3**.

Currently, data for 2010 to 2022 are included. The platform will continue to be updated with additional data as data become available.

Note that data presented in the portal may not perfectly match values that the BCHC members post in their reports and webpages. Differences could be due to a number of things such as differences in population denominators, age-standardization, stratified/sub-group values, etc.

GEOGRAPHIC REPRESENTATION

Effort has been made to provide city level data for the metrics. Some metrics are at the county level due to unavailability of city level data. This is noted in the metric subtitles or footnoted in the data displays. The county proxy was selected because it holds the largest share of the city population (per weighted place to county allocation factor published by the Missouri Census Data Center of the University of Missouri[1]). **Table 1** shows all the county proxies used.

Table 1: County proxies used for members when city level data were unavailable.

City	County Proxy
Austin, TX	Travis County, TX
Baltimore, MD	City has same boundaries as Baltimore City County, MD
Boston, MA	Suffolk County, MA
Charlotte, NC	Mecklenburg County, NC
Chicago, IL	Cook County, IL
Cleveland, OH	Cuyahoga County, OH
Columbus, OH	Franklin County, OH
Dallas, TX	Dallas County, TX
Denver, CO	City has same boundaries as Denver County, CO
Detroit, MI	Wayne County, MI
Fort Worth, TX	Tarrant County, TX
Houston, TX	Harris County, TX
Indianapolis, IN	Marion County, IN
Kansas City, MO	Jackson County, MO
Las Vegas, NV	Clark County, NV
Long Beach, CA	Los Angeles County, CA
Los Angeles, CA	Los Angeles County, CA
Miami, FL	Miami-Dade County, FL
Minneapolis, MN	Hennepin County, MN
New York City, NY	City has same boundaries as combination of 5 counties: Kings County, Queens County, New York County, Bronx County, Richmond County, NY
Oakland, CA	Alameda County, CA
Philadelphia, PA	City has same boundaries as Philadelphia County, PA
Phoenix, AZ	Maricopa County, AZ
Portland, OR	Multnomah County, OR
San Antonio, TX	Bexar County, TX
San Diego, CA	San Diego County, CA
San Francisco, CA	City has same boundaries as San Francisco County, CA

City	County Proxy
San Jose, CA	Santa Clara County, CA
Seattle, WA	King County, WA
Washington, DC	The city is synonymous with District of Columbia County, DC

CITY CLASSIFICATION

One of the unique features of this data platform is the ability to compare different metrics between cities. To put the differences between the cities into perspective, the users can group similar cities. Users have 2 options to group cities in the COMPARE CITIES page, a) "Group Cities By", b) "5 Comparison Cities".

GROUP CITIES BY

Using this function, users can stratify cities based on certain city characteristics when they view different metrics. This allows them to understand the underlying factors driving the differences between cites. Generally, the groups divide the cities into top one fourth and bottom three fourths for the characteristics (ration of 9:26).

The characteristics used are--

a) Region- census bureau assigned regions: West, South, Midwest, Northeast
b) Poverty- Poorest cities (18%+ poor)/ Less poor cities (<18% poor)
c) Population size- Largest (>1.3 million)/ Smaller (<1.3 million)
d) Population density- Highest pop. density (>10k per sq mi)/ Lower pop. density (<10k per sq mi)
e) Racial residential segregation- using dissimilarity index: Highly Segregated (50%+)/ Less Segregated (<50%)

5 COMPARISON CITIES

Another option available to users in Compare Cities is "5 Comparison Cities". This feature is especially useful for viewing racial inequities and trends over time. When users select fewer than 6 cities, options appear for visualizing time trends and race-stratification (if the chosen metric has race data).

As a starting point, "5 Comparison Cities" were chosen using the method described below. We would be happy to alter the comparison list based on user's local knowledge/expertise. Please offer feedback via our short web form so that we follow-up up with you. Note that the universe of comparable cities is limited to cities who are members of the BCHC.

The comparison city grouping was created in 3 steps--

Step 1: K-Means Clustering was used to classify the cities into groups based on underlying similarities in city demographic and health characteristics. Peer cities assignment used the following characteristics (standardized to their mean): population size, %poverty, %Hispanic, %pon-Hispanic Black, geographic latitude, geographic longitude, and death rate.

We used 2 outputs from K-Means Clustering. The first including population size and the second without.

***Step 2*:** We then looked at the concordance of the 2 outputs created in the previous step and grouped all cities into approximately 4 large groups.

Step 3: Next, we selected 5 peers for each city that was based mostly on membership within the groups determined in the prior stage and using the following criteria.

- Include at least one comparison city in the index city's region. Regional proximity was forced because members told us they often compare their jurisdiction to cities in their region (even if characteristics do not closely match).
- Local pre-conceptions about peer cities. This information was obtained from a variety of sources such as: a city's own reports where they compared themselves to other large cities (these reports tended to originate within a city's urban planning unit/development unit); and local knowledge from city stakeholders which included consulting with the BCHC members who requested that we add the '5 comparison cities' feature.

YEARS OF DATA AVAILABLE

The years for which data were available for the metrics from each source is given in Appendix 1. Where missing, data were imputed from the nearest available year to facilitate visualization and comparison across different metrics.

MISSING DATA, DATA CALCULATIONS, METRIC SUB-GROUPS

MISSING DATA

In the platform, all metrics are not available for all cities, for all years, and for all subgroups. The reason for unavailability varies; in some cases, the data source lacked data for a particular city or for a particular subgroup; in other cases, the metric-city data existed but data were censored if the number of cases were too small to report. When data are unavailable, this is notated either as an empty column in the graph, a label on the graph (data unavailable, "N/A"), or the pull-down menu category is grey and not able to be selected.

DATA CALCULATIONS

Detailed methodology for calculation of metrics is given in the next section ('Explanation of Metrics'). Communicable disease indicators are reported using crude rates. Mortality rates are age-adjusted to compare relative mortality risks among cities, different demographic groups, and over time. Standardization was done using the 2000 U.S. standard million population. Metric names and subtitles reflect differences regarding age-adjustment and/or crude rates.

METRIC SUB-GROUPS

Sex

Where available, data for sex categories are reported for male and female.

Race and Ethnicity

Categorization of race and ethnicity were limited by the data source. For most metrics, the default options were Hispanic, White non-Hispanic, Black non-Hispanic, and Asian non-Hispanic. For the metrics that are sourced from the U.S. Census Bureau's American Community Survey, race is classified as Hispanic, White non-Hispanic, Black Hispanic and non-Hispanic, Asian Hispanic and non-Hispanic. For the metrics that are sourced from the National Vital Statistics System (NVSS) of the National Center for Health Statistics, race is classified as Hispanic, White non-Hispanic, Black Hispanic and non-Hispanic, Asian/Pacific Islander Hispanic and non-Hispanic.

EXPLANATION OF METRICS

Each of the metrics and associated methodology are further explained below.

ACCESS TO HEALTH SERVICES

Health Insurance

Uninsured, All Ages and **Uninsured, Child** were calculated using the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates and is calculated for all civilian non-institutionalized population from table B27010. The metric 'Uninsured, All Ages' is population of all ages who don't have any private or public insurance and the denominator is total population. The metric 'Uninsured, Child', is the population less than 18 years who don't have any private or public insurance and the denominator is population who are less than 18 years.

Oral Health

Dental Care is reported as received from the Centers for Disease Control and Prevention's Population Level Analysis and Community Estimates (CDC-PLACES) and are based primarily on the Behavioral Risk Factor Surveillance System. The numerator is the respondents aged 18+ who visited the dentist or a dental clinic in the previous year and the denominator is the total respondents. For more details, please see Appendix 2.

Births

Prenatal Care was calculated using the Natality files from the National Vital Statistics System (NVSS) of the National Center for Health Statistics. For details, please refer to Appendix 2. The numerator is the number of live births where prenatal care began by the third month of pregnancy and the denominator is the total number of live births.

CHRONIC HEALTH CONDITIONS

Physical Activity

Adult Physical Inactivity is reported as received from the Centers for Disease Control and Prevention's Population Level Analysis and Community Estimates (CDC-PLACES) and are based primarily on the Behavioral Risk Factor Surveillance System. The BRFSS question used is “During the past month, other than your regular job, did you participate in any physical activities or exercises such as running, calisthenics, golf, gardening, or walking for exercise?”. The numerator is respondents aged 18+ who answered ‘no’ to the question and the denominator is all respondents. For more details, please see Appendix 2.

Teen Physical Inactivity, Teen Physical Activity Levels, Teen Physical Education, Teen Computer Time, and Teen TV Time were reported as received from the Youth Risk Behavior Survey (YRBS). The metric ‘Teen Physical Inactivity’ is the percent of respondents who were not physically active for at least 60 minutes even 1 day in the 7 days before the survey. ‘Teen Physical Activity Levels’ is the percent of respondents who did not participate in at least 60 minutes of physical activity per day for at least 5 days in the 7 days before the survey i.e., did not meet the physical activity recommendations set by the CDC. ‘Teen Physical Education’ is the percent of high school students who did not attend PE class 1 or more days in an average week when they were in school. ‘Teen Computer Time’ is the percent of respondents who played video or computer games or used a computer three or more hours per day on an average school day. ‘Teen TV Time’ is the percent of high school students who watched television for 3 or more hours per day on an average school day. The denominator for all metrics is all respondents. For more details, please see Appendix 2.

Respiratory Disease

Teen Asthma is reported as received from the Youth Risk Behavior Survey (YRBS) and is the percent of all respondents who were ever told by a doctor or nurse that they had asthma. For more details, please see Appendix 2.

Cancer

All Cancer Deaths, Lung Cancer Deaths, Breast Cancer Deaths, Colorectal Cancer Deaths, and Prostate Cancer Deaths were calculated using the Multiple Cause-of-Death mortality data from the National Vital Statistics System (NVSS) of the National Center for Health Statistics. For details, please see Appendix 2 and detailed ICD 10 codes used are given in Table 6.

Cardiovascular Disease

Cardiovascular Disease Deaths and **Heart Disease Deaths** were calculated the Multiple Cause-of-Death mortality data from the National Vital Statistics System (NVSS) of the National Center for Health Statistics. For details, please see Appendix 2 and detailed ICD 10 codes are given in Table 6.

High Blood Pressure prevalence is reported as received from the Centers for Disease Control and Prevention's Population Level Analysis and Community Estimates (CDC-PLACES) and is based primarily on the Behavioral Risk Factor Surveillance System. The numerator is respondents aged 18+ years who report ever having been told by a doctor, nurse, or other health professional that they have high blood pressure. Women who were told they have high blood pressure only during pregnancy and those who were told they had borderline hypertension were not included. The denominator is all respondents aged 18+ years. For more details, please see Appendix 2.

Blood Pressure Medication is reported as received from the Centers for Disease Control and Prevention's Population Level Analysis and Community Estimates (CDC-PLACES) based primarily on the Behavioral Risk Factor Surveillance System. The numerator is respondents aged 18+ years who reported taking medicine for high blood pressure. The denominator is respondents aged 18+ years who report having been told by a doctor, nurse, or other health professional they had high blood pressure other than during pregnancy (excluding those who refused to answer, had a missing answer, or answered “don’t know/not sure”). For more details, please see Appendix 2.

Diabetes and Obesity

Diabetes Deaths were calculated the Multiple Cause-of-Death mortality data from the National Vital Statistics System (NVSS) of the National Center for Health Statistics. For details, please see Appendix 2 and detailed ICD 10 codes are given in Table 6.

Diabetes prevalence is reported as received from the Centers for Disease Control and Prevention's Population Level Analysis and Community Estimates (CDC-PLACES) based primarily on the Behavioral Risk Factor Surveillance System. The percentage of diagnosed diabetes were number of respondents who reported ever having been told by a doctor, nurse, or other health professional that they have diabetes. The denominator is all respondents aged 18+ years. Diabetes during pregnancy was excluded. Crude prevalence is reported.

Adult Obesity is reported as received from the Centers for Disease Control and Prevention's Population Level Analysis and Community Estimates (CDC-PLACES) based primarily on the Behavioral Risk Factor Surveillance System. It is the percentage of the population aged 18+ years who had body mass index (BMI) of 30 or above. The BMI was derived from self-reported height and weight. Respondents who reported being pregnant, extremes of height (<3 ft or ≥8 ft), extremes of

weight (<50 lbs. or ≥650 lbs.) or if height or weight data are missing were removed from the analysis.

Teen Obesity is reported as received from the Youth Risk Behavior Survey (YRBS). Obesity for children is defined as BMI at or above the 95th percentile of children of the same age and sex. The denominator is all respondents. For more details, please see Appendix 2.

Disability

People with Disabilities was calculated using the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates and is calculated from table B18101. The numerator is population of all ages who reported having any long-term disability and the denominator is all civilian noninstitutionalized population.

Dietary Quality

Teen Soda and **Teen Breakfast** data are reported as received from the Youth Risk Behavior Survey (YRBS). 'Teen Soda' is the percent who responded 'yes' to the YRBS question "Whether they drank a can, bottle, or glass of soda or pop such as Coke, Pepsi, or Sprite, not counting diet soda or diet pop one or more times per day, during the 7 days before the survey". 'Teen Breakfast' is percent who reported eating no breakfast during the 7 days before the survey. The denominator is all respondents. For more details, please see Appendix 2.

INFECTIOUS DISEASES

Respiratory Infection

Pneumonia or Influenza Deaths was calculated using the Multiple Cause-of-Death mortality data from the National Vital Statistics System (NVSS) of the National Center for Health Statistics. For details, please see Appendix 2 and detailed ICD 10 codes are given in Table 6.

New Tuberculosis Cases are reported as received from the National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention Atlas Plus of the Centers for Disease Control and Prevention**.** Tuberculosis Cases which met the clinical case classification or laboratory confirmed are counted. Crude rates per 100,000 population were calculated for adults aged 18+.

Flu Vaccine, Adult is reported as received from the Centers for Medicare and Medicaid Services (CMMS), of the U.S. Department of Health and Human Services. The denominator represents all Medicare beneficiaries enrolled in the fee-for-service (FFS) program. The numerator represents the subset of beneficiaries who received a flu vaccine in the year specified; CMMS derived this from administrative claims for flu vaccination.

COVID-19 Deaths was calculated using the Multiple Cause-of-Death mortality data from the National Vital Statistics System (NVSS) of the National Center for Health Statistics. Death rates are reported for single years. For details, please see Appendix 2 and detailed ICD 10 codes are given in Table 6.

COVID-19 Vaccinations are reported as received from the COVID Data Tracker historic county-level data, published by the Centers for Disease Control and Prevention (CDC). More information about these data are available at the COVID Data Tracker website. The CDC compiled data from its vaccine reporting partners including jurisdictional partner clinics, retail pharmacies, long-term care facilities, dialysis centers, Federal Emergency Management Agency and Health Resources and Services Administration partner sites, and federal entity facilities. Completion of the COVID-19 primary vaccine series means that an individual has had a second dose of a two-dose vaccine or one dose of a single-dose vaccine. Completion is assessed cumulatively starting when the vaccines were first available (2020) through the end of the reporting year for the data in the BCHI platform. For the metric 'COVID-19 Vaccination, Adults', the numerator is the number aged 18+ years who completed the COVID-19 primary vaccine series. The denominator is the total population aged 18+. For the metric 'COVID-19 Vaccination, Seniors', the numerator is the number aged 65+ years who completed the COVID-19 primary vaccine series. The denominator is the total population aged 65+.

Sexually Transmitted Disease

HIV/AIDS Prevalence, New Chlamydia Cases, Syphilis Prevalence (primary & secondary syphilis), **Syphilis in Newborns** (congenital syphilis), and **New Gonorrhea Cases** are reported as received from the National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention Atlas Plus of the Centers for Disease Control and Prevention. Crude rates per 100,000 population were calculated for each STD (except congenital syphilis). Each rate was calculated by dividing the total incidence or prevalence for the calendar year by the population aged 13+ years for that calendar year and then multiplying the result by 100,000. Congenital syphilis is reported per 100,000 live births.

HIV-Related Deaths was calculated using the Multiple Cause-of-Death mortality data from the National Vital Statistics System (NVSS) of the National Center for Health Statistics. For details, please see Appendix 2 and detailed ICD 10 codes are given in Table 6.

LIFE EXPECTANCY AND DEATHS

Deaths

Deaths from All Causes was calculated using the Multiple Cause-of-Death mortality data from the National Vital Statistics System (NVSS) of the National Center for Health Statistics. For details, please see Appendix 2 and detailed ICD 10 codes are given in Table 6.

Premature Death (age < 75) was calculated using the Multiple Cause-of-Death mortality data from the National Vital Statistics System (NVSS) of the National Center for Health Statistics. 'Premature Death (age < 75)' is defined as years of potential life lost from all causes of death before age 75. It is calculated using Dranger and Remington's[2] method and is age-adjusted and reported per 100,000 population. Weights for age-adjusting premature deaths were calculated using the US 2000 standardized population but were adjusted to include only the population aged 74 and younger. This is calculated as three-year moving averages. For details on years aggregated, please see Appendix 2.

Life Expectancy at Birth

Life Expectancy was calculated as the life expectancy at birth using the Multiple Cause-of-Death mortality data from the National Vital Statistics System (NVSS) of the National Center for Health Statistics. Life expectancies were calculated as 2-year moving averages (**Table 2**, shows how the data were aggregated for each period). The death incidences were calculated for all members at city level except for Las Vegas, NV and Louisville, KY where county level data were used. The population counts were obtained from ACS-1-year estimates. Abridged life tables and life expectancies were calculated using R Software Demo Tools package. To overcome the problem of unavailability of the population data for under 1-year old by race and sex strata, population 'graduation' into single age groups was done for race-sex specific population using the formula by Sprague[3] and implemented using the R package 'Demo Tools'. Life expectancy estimates based on less than 300 deaths were suppressed.

Table 2: The years aggregated for the numerators and the denominators for Life Expectancy

"year" label	Numerator: Deaths		Denominator: Population	
2011	2010	2011	2010	2011
2012	2011	2012	2011	2012
2013	2012	2013	2012	2013
2014	2013	2014	2013	2014
2015	2014	2015	2014	2015
2016	2015	2016	2015	2016
2017	2016	2017	2016	2017
2018	2017	2018	2017	2018
2019	2018	2019	2018	2019

2020	2019	2019*	2019	2019*
2021	2019*	2021	2019*	2021
2022	2021	2022	2021	2022

2019*- Population estimates in the 2020 ACS-1-year surveys were unreliable. 2019 ACS-1-year population estimates were used as proxy in accordance with the recommendations by the Census Bureau.

MATERNAL AND CHILD HEALTH

Births

Teen Births, Low Birthweight, and **Preterm Births** were calculated using the Natality files from the National Vital Statistics System (NVSS) of the National Center for Health Statistics. For details, please refer to Appendix 2. For the metric 'Teen Births', the numerator is the number of live births to mothers aged 15 to 19 years and denominator is the total female population aged 15 to 19 from the American Community Survey 1-year population estimates. 'Low Birthweight' is defined as the percentage of babies born under 2,500 grams. The denominator is the total number of live births. 'Preterm Births' is defined as percentage of live births delivered at less than 37 completed weeks of gestation (gestational age calculated by obstetric estimation).

Deaths

Infant Deaths and **Maternal Deaths** were calculated using the Multiple Cause-of-Death mortality data from the National Vital Statistics System (NVSS) of the National Center for Health Statistics. For details, please see Appendix 2 and detailed ICD 10 codes are given in Table 6. 'Infant Deaths' is defined as death within 1 year of life. It's expressed per 1,000 live births. 'Maternal Deaths' is defined as maternal death during pregnancy or within 42 days of pregnancy termination due to complications of pregnancy, childbirth, and puerperium and not due to accidental or incidental causes. It's expressed per 100,000 live births.

Birth Control

Teen Birth Control were reported as received from the Youth Risk Behavior Survey (YRBS) and is the percent of teens who did not use any method to prevent pregnancy during last sexual intercourse among students who were currently sexually active. For more details, please see Appendix 2.

MENTAL HEALTH AND SUBSTANCE USE

Mental Health

Adult Mental Distress is reported as received from the Centers for Disease Control and Prevention's Population Level Analysis and Community Estimates (CDC-PLACES) based primarily on the Behavioral Risk Factor Surveillance System. The

numerator is respondents aged 18+ years who report their mental health being "not good" for 14 or more days during the past 30 days. The denominator is all respondents aged 18+ years. For more details, please see Appendix 2.

Suicide mortality rates were calculated using the Multiple Cause-of-Death mortality data from the National Vital Statistics System (NVSS) of the National Center for Health Statistics. For details, please see Appendix 2 and detailed ICD 10 codes are given in Table 6.

Electronic Bullying, **School Bullying**, **Teen Mental Distress** and **Teen Suicidal Ideation** were reported as received from the Youth Risk Behavior Survey (YRBS). 'Electronic Bullying' is the percent of high school students who recounted being bullied through texting, Instagram, Facebook, or other social media, during the 12 months before the survey. 'School Bullying' is the percent of high school students who recounted being bullied on school property in the past year. 'Teen Mental Distress' is the percent of high school students who felt sad or hopeless almost every day for 2 weeks or more in a row so that they stopped doing some usual activities, during the 12 months before the survey. 'Teen Suicidal Ideation' is the percent of high school students who planned about how they would attempt suicide during the 12 months before the survey. The denominator for all metrics were all respondents. For more details, please see Appendix 2.

Substance Use

Opioid Overdose Deaths was calculated using the Multiple Cause-of-Death mortality data from the National Vital Statistics System (NVSS) of the National Center for Health Statistics. For details, please see Appendix 2 and detailed ICD 10 codes are given in Table 6.

Adult Binge Drinking and **Adult Smoking** were reported as received from the Centers for Disease Control and Prevention's Population Level Analysis and Community Estimates (CDC-PLACES) based primarily on the Behavioral Risk Factor Surveillance System. The percentage of adults who binge drank is based on the Behavioral Risk Factor Surveillance System question about how many drinks a person had on one occasion in the past 30 days. Women who answered "four" and men who answered "five" are considered binge drinkers. The denominator is adults aged 18+ years who report having a specific number (including zero) of drinks on an occasion in the past 30 days (excluding those who refused to answer, had a missing answer, or answered "don't know/not sure"). 'Adult Smoking' is the percent of respondents aged 18+ years who report having smoked ≥100 cigarettes in their lifetime and currently smoke every day or some days. The denominator is adults aged 18+ years excluding respondents who refused to answer. For more details, please see Appendix 2.

Teen Alcohol, **Teen Smoking**, and **Teen Marijuana** were reported as received from the Youth Risk Behavior Survey (YRBS). 'Teen Alcohol' is the percent of high school

students who currently drank at least one drink of alcohol, on at least 1 day during the 30 days before the survey. 'Teen Smoking' is the percent of high school students who currently reported smoking cigarettes on at least 1 day during the 30 days before the survey. 'Teen Marijuana' is the percent of high school students who currently used marijuana also called grass, pot, or weed, one or more times during the 30 days before the survey. The denominator for all metrics were all respondents. For more details, please see Appendix 2.

CLIMATE AND BUILT ENVIRONMENT

Air Pollution

Poor Air Quality and **Hazardous Air Quality** are classified using the Air Quality Index (AQI) values published by the Environmental Protection Agency. EPA derives AQI for four major air pollutants regulated by the Clean Air Act: ground level ozone, particle pollution, carbon monoxide, and sulfur dioxide. EPA considers AQI values below 50 as 'good air quality' with little or no potential to affect public health; whereas days with AQI values >50 may be of health concern for sensitive groups, and days with AQI >100 are considered unhealthy for all people. The platform metric numerator is the number of days when AQI was >50 (for 'Poor Air Quality') or >100 (for 'Hazardous Air Quality') and the denominator is number of days for which AQI is recorded.

Food Access

Limited Supermarket Access is calculated from indicators in the Food Access Research Atlas of the U.S. Department of Agriculture. Food Access Research Atlas includes several indicators to measure food access at the census tract level. We aggregated these data to the city level using the population weighted census tract to census places (city) allocation factor published by the Missouri Census Data Centre of the University of Missouri[1]. For 'Limited Supermarket Access', the numerator is the number of low-income people who did not live within half mile of supermarket and the denominator is the total low-income population. Low-income is defined as people with income below 200% of Federal Poverty level.

Park Access

Green Space Access, Investment in Parks and **City Park System** as measured by ParkScore® were reported as received from the ParkScore® database of the Trust for Public Land.

The Trust for Public Land developed the ParkScore® index for the 100 largest cities in the US, to measure how well cities are meeting the need for parks. Cities are awarded points based on analysis of four important characteristics of an effective park system: acreage, investment, amenities, and access. The metric 'Green Space

Access' is percent of residents living within a 10-minute walk of green spaces which include: publicly owned local, state, and national parks, trails, and open space; school parks with a joint-use agreement with the local government; and privately owned parks that are managed for full public use. 'Investment in Park' is based on total spending per resident and is the sum of all public spending, nonprofit spending, and volunteer hours. 'City Park System' is the ParkScore® Index scaled from 0 – 100 with higher score representing better performance.

Active Transportation

Bikeability (measured as Bike Score®) and **Walkability** (measured as Walk Score®) were reported as received from the Redfin Corporation. Bike Score® is a patented measure of whether an area is good for biking. For a given location, a Bike Score® is calculated by measuring bike infrastructure (lanes, trails, etc.), hills, destinations and road connectivity, and the number of bike commuters. Walk Score® measures the walkability of any address using a patented system. For each address, Walk Score® analyzes hundreds of walking routes to nearby amenities. Points are awarded based on the distance to amenities in each category. Walk Score® also measures pedestrian friendliness by analyzing population density and road metrics such as block length and intersection density.

Walking to Work and **Riding Bike to Work** were calculated using the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates (from table B08137). The numerator is those who reported walking or riding a bicycle to work respectively. The denominator is all workers aged 16+ years who did not work at home.

Transportation

Public Transit Access (measured as Transit Score®) is reported as received from the Redfin Corporation. Transit Score® is a patented measure of how well a location is served by public transit. Transit Score is based on data released in a standard format by public transit agencies. To calculate a Transit Score®, a "usefulness" value to nearby transit routes based on the frequency, type of route (rail, bus, etc.), and distance to the nearest stop on the route is assigned. The "usefulness" of all nearby routes is summed up and normalized to a score between 0-100.

Lack of Car was calculated using the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates and is calculated from table B25044. The numerator is the number of housing units (both owner and renter occupied) with no vehicles available. The denominator is all occupied housing units.

Public Transportation Use and **Drives Alone to Work** were calculated using the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates and was calculated from table B08137. The numerator is those who reported using public

transport, or those who drive to work alone in a car truck, or van respectively. The denominator is all workers aged 16+ years who did not work at home.

Longer Driving Commute Time was calculated using the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates and is calculated from table B08012. The numerator is those who reported commute time to work to be more than 30 minutes. The denominator is all workers aged 16+ years who did not work at home.

Heat and Disasters

Number of Heat Waves was calculated using the data from the Global Historic Climatic Network- daily database**.** It is defined as the average number of heat waves per year during the period 2010-2022. For details on data source and calculations, please see Appendix 2.

Longer Summer was calculated using data from the Global Historic Climatic Network- daily database. The metric approximates the increase in the number of summer days the city has been experiencing in recent years compared to the number of summer days the city experienced 60 years ago. For details on data source and calculations, please see Appendix 2.

Climate Disasters was derived from the Federal Emergency Management Agency's (FEMA), Disaster Declaration Database (National Emergency Management Information System). We derived a metric to represent the average number of days per year during 2001-2022 that the city was affected by a federally declared climate-related disaster; this included severe storms, hurricanes, floods, wildfire, tornadoes, etc. but not heat waves. Federal disaster declarations were declared for the state or county either because (a) a catastrophic event occurred within the county or (b) it occurred outside of the county but within the county's disaster response region and the county needed to provide a high level of aid to the disaster response. For details on data source and calculations, please see Appendix 2.

Community Social Vulnerability to Climate Disaster is the percent of tracts in the city that were in the top quartile of Social Vulnerability Index (SVI) score (0.75) for population in the US.

The SVI was created by the Agency for Toxic Substance and Disease Registry's (ATSDR) Geospatial Research, Analysis, & Services Program (GRASP) to help public health officials and emergency response planners identify and map the communities that will most likely need support before, during, and after a hazardous event. SVI indicates the relative vulnerability of every U.S. Census tract using 16 social factors that are grouped into 4 themes. For detailed methodology please refer to documentation from the ATSDR https://www.atsdr.cdc.gov/placeandhealth/svi/documentation/pdf/SVI2020Documentation_08.05.22.pdf.

The 4 themes and 16 factors are:

Socioeconomic Status: Below 150% Poverty, Unemployed, Housing cost burden, No high school diploma, No health insurance.

Household Characteristics: Age 65+, Age<18, civilian with disability, Single-parent household, English language proficiency.

Racial & Ethnic Minority Status: Hispanic/Latino (of any race), Non-Hispanic Black, NH Asian, NH American Indian, or Alaska Native, Non-Hispanic Native Hawaiian or Pacific Islander, Non-Hispanic Two or more races, Non-Hispanic Other races.

Housing type and Transportation: Multi-unit structure, mobile homes, crowding, no vehicle, group quarters.

Lead Poisoning

Child Lead Levels use two separate cut-offs: **5 mcg/dL** (micrograms of lead per deciliter of whole blood) and **10 mcg/dL**. The numerator is the number of children with elevated blood levels and the denominator is the number of children under 6 who were screened.

In the BCHI we report (as received) the county data are available from Childhood Lead Poisoning Prevention Program (CLPPP) of the Centers for Disease Control and Prevention (CDC) and augment this with county CLPPP data reported by the state of Texas, state of California, and New York City.

Child Lead Testing is the number of children under 6 out of whole population under 6 who were tested for elevated blood lead level. The population data is obtained from the annual intercensal estimates for the most recent U.S. Census data. This metric is included to aid in the interpretation of the result of elevated blood lead level since screening is not mandated for children under 6 in all cities. The volume of children screened affects the percent of children detected with elevated blood lead level.

Background on childhood lead level monitoring and reporting over time.

CDC has noted that there is no safe childhood blood lead level and over the past 10 years has recommended testing, reporting, and action based on increasingly lower values of lead in the blood. Changes in blood lead surveillance are important for protecting the public's health but create challenges for surveillance systems and reporting.

Prior to 2012, children were identified as having a blood lead "level of concern" if the test result was >=10 mcg/dL. In 2012, CDC introduced the concept of a blood lead "reference value" (BLRV) to identify children with higher levels of lead in their blood compared to most children. The BLRV for

children corresponding to the 97.5th percentile was established at 5 mcg/dL in 2012, and then lowered to 3.5 mcg/dL in 2021.

The BLRV is not a clinical reference level defining an acceptable range of blood lead levels in children nor is it a health-based toxicity threshold. The BLRV is based on the 97.5th percentile of the blood lead distribution in U.S. children ages 1–5 years from the most recent two cycles of data from the National Health and Nutrition Examination Survey. For that reason, the BLRV changes when new NHANES data are released. See CDC's website for more details https://www.cdc.gov/nceh/lead/data/blood-lead-reference-value.htm

Housing Lead Risk was calculated using the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates and is calculated from table B25034. The numerator is the housing units built before 1950. The denominator is all housing units.

DEMOGRAPHICS

Population

Single-parent Families was calculated using the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates and is calculated from table B11005. The numerator is the number of households with a child below 18 with female only or male only householder (either family or non-family households). The denominator is all households.

Population Density per square mile was calculated using the weighted population from the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates (table B01003) and the land area is acquired from the gazetteer files published by the U.S. Census Bureau. The numerator is the total population of the city, and the denominator is the land area of the city.

Seniors and **Children** were calculated using the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates and is calculated for total population. These are calculated using table B01001 which gives the population counts by sex and age groups. The metric 'Seniors' is the percent of population who are aged 65+ years. The metric 'Children' is the percent of population aged less than 5 years.

Race/Ethnicity

Minority Population was calculated using the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates and is calculated from table B03002.

This was calculated by subtracting the percentage of people who reported being non-Hispanic white from 100. The denominator is total population.

Language and Nativity

Primarily Speak English, Primarily Speak Chinese, and **Primarily Speak Spanish** were calculated using the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates and is calculated from table C16001. The numerators are the people who reported speaking English, Spanish or Chinese primarily at home respectively. The denominator is population aged 5 years and older.

Foreign Born Population was calculated using the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates and is calculated from table B05003. The numerator is all people who reported being born in a foreign country. The denominator is the total population.

SOCIAL AND ECONOMIC FACTORS

Housing

Vacant Housing Units, **Owner Occupied Housing**, **Renters vs. Owners** were calculated using the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates and is calculated for housing units. 'Vacant Housing Units' is the percent of total housing units that are vacant (table B25002). 'Owner Occupied Housing' is the percent of occupied housing units that are occupied by the owner (table B25003). The metric 'Renters vs. Owners' is the ratio of renter occupied to owner occupied housing units among all occupied housing units (table B25003).

Homelessness metrics were calculated using data from the Point-in-Time (PIT) survey conducted by the Continuum of Care (CoC) Homeless Assistance Programs of the U.S. Department of Housing and Urban Development (HUD). The PIT is an unduplicated 1-night estimate (usually done in January) of both 'sheltered' and 'unsheltered' homeless populations, conducted according to HUD protocols. 'Sheltered' refers to people who are staying in emergency shelters or transitional housing programs. 'Unsheltered' refers to people whose primary nighttime residence is a public or private place not designated for, or ordinarily used as, a regular sleeping accommodation for people (for example, the streets, vehicles, or parks).

For the metric 'Homeless, Total', the numerator is total unhoused population. The denominator is the total population of the city (table B01003) from the U.S Census Bureau's American Community Survey (ACS) 5-year estimates. For the metric 'Homeless, Children' the numerator is all unhoused population who were less than 18 years old. The denominator is total unhoused population. For the metric 'Vacant Housing and Homelessness', the numerator is total number of vacant rental units (table B25004) from the U.S Census Bureau's American Community Survey (ACS) 5-year estimates. The denominator is all unhoused population.

Education

Preschool Enrollment was calculated using the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates and is calculated for population over 3 years of age from table B14003. The numerator is 3- and 4-year-olds enrolled in school (both public and private schools), the denominator is the total population of 3- and 4-year-old age group.

College Graduates was calculated using the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates and is calculated from table B15002. The numerator is the people who reported having educational attainment of bachelor's degree or higher. The denominator is population over 25 years of age.

Income

Poverty in All Ages and **Poverty in Children** were calculated using the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates and is calculated for population for whom the poverty status was determined from table B17001. For the metric 'Poverty in All Ages', the numerator is population of all ages who reported family income for the past 12 months to be below the federal poverty level. The denominator is total population for whom the poverty status was determined. For the metric 'Poverty in Children', the numerator is population below 18 years of age who reported family income in the past 12 months to be below poverty level and the denominator is population below 18 years for whom the poverty status was determined.

Poverty and Near Poverty in All Ages was calculated using the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates and is calculated for population for whom the poverty status was determined (table C17002). The numerator is all whose ratio of income to federal poverty level was less than 200%.

Per-capita Household Income is reported from table B19301 of the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates. This is reported for the total population.

Households with Higher-Incomes were calculated using the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates and is calculated from table B19001. The numerator is households whose reported income in the past year was greater than $50,000. The denominator is all households.

Income-related

Public Assistance was calculated using the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates and is calculated from table B19058. The numerator is households who received cash public assistance or Food Stamps/SNAP. The denominator is all households.

Unemployment was calculated using the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates and is calculated from table B23001. The numerator is people in civilian workforce who reported being currently unemployed. The denominator is civilians 16+ years of age.

Service Workers was calculated using the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates and is calculated from table C24010. The numerator is people who reported being employed in service or labor occupations. The denominator is civilians in workforce aged 16+ years.

Excessive Housing Cost was calculated using the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates and is calculated from tables B25074 and B25091. The numerator is all renters who reported the rent, or owner-occupied housing units who reported the mortgage to be >35% of the household income. The denominator is all housing units.

Income Inequality

Household Income Inequality expressed by Index of Concentration at the Extremes (ICE) was calculated using the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates and is calculated from table B19001. The numerator is households with incomes at the extremes of the national income distribution namely those with annual income < $25,000 (bottom 20%) or > $100,000 (top 20%). The denominator is all households.

Income Inequality expressed as GINI index is reported from table B19083 of the U.S. Census Bureau's American Community Survey (ACS) 5-year estimates. This is reported for all households.

Racial segregation indices

Racial Segregation, White and Non-White; Racial Segregation, White and Black; Racial Segregation, White and Asian; and **Racial Segregation, White and Hispanic** were assessed using the index of dissimilarity. The metric reflects racial segregation across neighborhoods (census tracts) in the city (census places).

The source data are tract-level demographic data from the 5-year American Community Survey (ACS). Tracts were assigned to cities (census places) based on the population weighted allocation factor published by the Missouri Census Data Center of the University of Missouri[1]. ACS surveys used were for the time periods 2010-2014, and 2015-2019.

We chose to show segregation using the index of dissimilarity because it is among the most widely reported segregation metrics. The index reflects the evenness with which two groups (for example, Black and white residents) are distributed across the component geographic areas (census tracts) within a larger area (census places). The index ranges from 0 (complete integration) to 100 (complete

segregation). The index is smallest when majority and minority populations are evenly distributed.

Generally, segregation values of 50% or higher are considered 'high' segregation (see more brookings.edu/blog/the-avenue/2018/12/17/black-white-segregation-edges-downward-since-2000-census-shows/)

We include racial segregation in our platform because it is widely considered a fundamental cause of racial disparities in health in the United States. As described by Williams et al, and others, poor health is more common in places that are segregated from socioeconomic mobility and where there is concentrated exposure to social and physical hazards that harm health. CDC's Healthy People 2030 cites residential segregation as an example of how structural discrimination harms health (see their Social and Community Context domain).

As has been explained in numerous publications (academic and other publications), the persistence of high racial residential segregation rates across US cities is due to a long history and continuation of tactics that selectively restrict occupancy in neighborhoods reserved for whites. Tactics from the past and present include restricted covenants, redlining communities of color, intimidation, violence, and selective use of law enforcement and mass incarceration.

VIOLENCE AND INJURY

Crime Incidents

Violent Crime rates are reported as received from the Uniform Crime Reporting program of the Federal Bureau of Investigation (F.B.I.). Under the program all city law enforcement agencies submit offense data to the FBI. Violent crime incidences included are murder, aggravated assault, robbery, and rape. The population estimates were computed by the FBI based on the 2010 decennial population counts and the rate of growth for the individual city. Crude rate per 100,000 population is reported.

Homicides rates were calculated using the Multiple Cause-of-Death mortality data from the National Vital Statistics System (NVSS) of the National Center for Health Statistics. For details, please see Appendix 2 and detailed ICD 10 codes are given in Table 6.

Deaths

Injury Deaths was calculated using the Multiple Cause-of-Death mortality data from the National Vital Statistics System (NVSS) of the National Center for Health Statistics. For details, please see Appendix 2 and detailed ICD 10 codes are given in Table 6.

Firearm Deaths was calculated using the Multiple Cause-of-Death mortality data from the National Vital Statistics System (NVSS) of the National Center for Health Statistics. For details, please see Appendix 2 and detailed ICD 10 codes are given in Table 6.

Police Killings and **Racial Disparity in Police Killings** were reported from the Mapping Police Violence database which records information on killings by police nationwide since 2013. This information has been sourced both from official police reports in some states (e.g., California, Texas etc.) and from the Fatal Encounters database, a nationwide impartial crowdsourced database on police killings. **Police Killing** is defined as a case where a person dies as a result of being shot, beaten, restrained, intentionally hit by a police vehicle, pepper sprayed, tasered, or otherwise harmed by police officers, whether on-duty or off-duty. Only killings where the person was unarmed have been considered. The numerator is the sum of all killings during the period and the denominator is the American Community Survey 5-year population estimates for 2018-2022. The rate is reported as crude annual rate per 1,000,000 population. **Racial Disparity in Police Killing** is calculated as the difference in annual police killing rate between unarmed Black people and unarmed white people.

Motor Vehicle Deaths was calculated using the Multiple Cause-of-Death mortality data from the National Vital Statistics System (NVSS) of the National Center for Health Statistics. For details, please see Appendix 2 and detailed ICD 10 codes are given in Table 6.

School Violence

Weapons in School is reported as received from the Youth Risk Behavior Survey (YRBS). It is the percentage of high school students who carried a weapon such as a gun, knife, or club, on at least 1 day during the 30 days before the survey. The denominator is all respondents. For more details, please see Appendix 2.

Fighting in School is reported as received from the Youth Risk Behavior Survey (YRBS). It is the percentage of high school students who were involved in a physical fight one or more times during the 12 months before the survey. The denominator is all respondents. For more details, please see Appendix 2.

GLOSSARY OF TERMS

The following terms are defined by the CDC's Principles of Epidemiology in Public Health Practice, Third Edition: An Introduction to Applied Epidemiology and Biostatistics and the United States Cancer Statistics.

Rate – A rate is a measure of the frequency with which an event occurs in a defined population over a specified period of time.

Crude rate – The total number of cases of a particular disease or condition over the total population size for a given period of time. As the crude rate is influenced by the underlying age distribution of the state's population, cities will often report an age-adjusted rate.

Age-adjusted rate – Uses a standard population, generally the 2000 U.S. standard population, that is based on that year's population age groups. Using direct standardization, these populations by age group serve as weights for calculating the age-adjusted rate. This ensures differences in rates are not due to different age distributions of the populations.

APPENDICES

Appendix 1. List of data sources

Table 3: Data aggregation and period for which data are available from each data source. The symbol ~ indicates the data are available for most cities.

Source	Multiple/ Single year	Time Period
American Community Survey, U.S. Census Bureau	Multiple	2008-2012, 2013-2017, 2018-2022
Childhood Lead Poisoning Prevention Program, CDC	Single	2012 to 2017~
COVID Data Tracker, CDC	Single	2021, 2022
Economic Research Service, U.S. Department of Agriculture	Multiple	2015, 2019
Mapping Police Violence	Multiple	2013-2022
National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention AtlasPlus, CDC	Single	2010 to 2021
National Vital Statistics System (NVSS), CDC	Multiple	2010 to 2022
ParkScore®, The Trust for Public Land	Single	2015 to 2022
Population Level Analysis and Community Estimates (CDC-PLACES)	Multiple	2014 to 2021
U.S. Environmental Protection Agency	Single	2010 to 2022
Uniform Crime Reporting, FBI	Single	2010 to 2022
Walk Score®, Redfin Corporation	Single	2022
Youth Risk Behavior Surveillance System, CDC	Single	2011, 2013, 2015, 2017, 2019~
Global Historical Climatology Network, National Centers for Environmental Information	Multiple	2010 -2022
ATSDR's Geospatial Research, Analysis and Services Program (GRASP), Centers for Disease Control and Prevention	Multiple	2010, 2014, 2016, 2018, 2020
Disaster Declarations Summaries, Federal Emergency Management Agency	Multiple	2001-2021

Appendix 2. Details on data sources.

AMERICA COMMUNITY SURVEY (ACS)

American Community Survey (ACS) 5-year estimates summary files were summarized at "summary level 160" for Places and analytic variables were created. Places FIPS codes, which are unique for city (7 digit- State FIPS+ Place FIPS), were used to isolate the data for the BCHC cites.

The 5-year estimates from the ACS are "period" estimates that represent data collected over a period of time. The primary advantage of using multiyear estimates is the increased statistical reliability of the data for less populated areas and small population subgroups. Margins of error were not considered in the calculations on our data platform. Three non-overlapping 5-year surveys were used. The data for the years 2010 to 2012 are from ACS 5-year survey 2008-2012. The data for the years 2013 to 2017 are

from ACS 5-year survey 2013-2017. The data for the years 2018 onwards are from ACS 5-year survey 2018-2022.

YOUTH RISK BEHAVIOR SURVEILLANCE SURVEY (YRBS)

Adolescent (teen) health metrics come from the high school Youth Risk Behavior Survey (YRBS). In general, the high-school YRBS are intended to be used to monitor health trends, identify emerging issues, and plan and evaluate programs that can help improve adolescent health. The high school survey is collected every two years (every odd numbered year, usually in spring semester) in grade 9–12 classrooms across the U.S.. The survey is self-administered anonymously and takes one class period (approximately 45 minutes) to complete. Survey questions have undergone test-retest analysis and demonstrated good reliability.

Our platform displays data from CDC's school-district level summary estimates of select survey questions. Overall estimates and estimates for sex and race/ethnic strata meet the standard of producing estimates within ±5% at a 95% confidence interval level. Our data platform aimed to show YRBS metrics spanning 2010—2021. However, in recent years states or school districts have not participated. By 2021, YRBS data are only available for approximately 35% of the BCHC cities and for only some of the metrics.

CENTERS FOR DISEASE CONTROL AND PREVENTION (CDC)-PLACES

The Centers for Disease Control and Prevention's Population Level Analysis and Community Estimates (CDC-PLACES) measures are based primarily on the Behavioral Risk Factor Surveillance System (BRFSS).

The BRFSS is designed to provide state-level estimates. In order to obtain stable city-level estimates, the CDC used a small-area-estimation modeling approach that incorporates individual-level attributes from the survey respondent and city-level census socio-demographic data into the estimation procedure. CDC's estimates utilize survey respondent age-sex-race demographics during the regression estimation procedure; for that reason, CDC-PLACES does not provide the metrics by subgroups of race or sex. More information is available here: www.cdc.gov/places/index.html

NATIONAL VITAL STATISTICS SYSTEM (NVSS)

General Notes:

Restricted-use micro-data files for Natality[4] and Multiple Causes-of-Death Mortality[5] from the NVSS were used for the calculation of these metrics. These files have city (census Place) FIPS codes identifier for residents of cities with population size ≥ 100,000.

Geography: Deaths are assigned to the reported city of residence of the deceased (based on city FIPS code). Births are assigned to the city of residence reported by the mother. Rates are reported for all cities at city (census Place) level except for Las Vegas, NV and

Louisville, KY. For these cities, the data are reported at county level for all metrics with the exception of Infant Deaths, Maternal Deaths, Prenatal Care Low Birth Weight, and Preterm Birth for which data are reported at city (census Place) level .

Categorizing race/ethnicity: Race/ethnicity categories selected were limited by the race categories available from the NVSS data and the population estimates from the U.S. Census Bureau's American Community Survey. Race classifications used are Hispanic, White non-Hispanic, Black Hispanic and non-Hispanic, Asian/Pacific Islander Hispanic and non-Hispanic. For mortality, the race is classified according to the race of the deceased recorded in the death certificate. For the natality measures the race is classified according to the race of the mother.

Years

The rates were calculated as three-year moving averages (except for 2010 and 2021 where two-year averages are used) for most metrics (exceptions are Infant Deaths and Maternal Deaths which are detailed below). The middle year is used as the representative year.

The **Table 4** & **5** shows the years aggregated for the numerators and the denominators.

Table 4: The years aggregated for the numerators and the denominators for mortality metrics.

"year" label	Numerator: Deaths or Births			Denominator: Population/Live Births		
2011	--	2010	2011	--	2010	2011
2012	2010	2011	2012	2010	2011	2012
2012	2011	2012	2013	2011	2012	2013
2013	2012	2013	2014	2012	2013	2014
2014	2013	2014	2015	2013	2014	2015
2015	2014	2015	2016	2014	2015	2016
2016	2015	2016	2017	2015	2016	2017
2017	2016	2017	2018	2016	2017	2018
2018	2017	2018	2019	2017	2018	2019
2019	2018	2019	2020	2018	2019	2019*
2020	2019	2020	2021	2019	2019*	2021
2021	2020	2021	2022	2019*	2021	2022
2022	2021	2022	--	2021	2022	--

2019*- Population estimates in the 2020 ACS-1-year surveys were unreliable. 2019 ACS-1-year population estimates were used as proxy in accordance with the recommendations by the Census Bureau.

For **COVID-19 Deaths,** rates are reported for individual years beginning in 2020.

For **Infant Deaths**, rates were calculated as five-year moving averages. The years aggregated are as follows. The end year is used as the representative year.

Table 5: The years aggregated for the numerators and the denominators for infant mortality.

"year" label	Numerator: Deaths					Denominator: Live Births				
	Mortality files years used					Natality files years used				
2014	2010	2011	2012	2013	2014	2010	2011	2012	2013	2014
2015	2011	2012	2013	2014	2015	2011	2012	2013	2014	2015
2016	2012	2013	2014	2015	2016	2012	2013	2014	2015	2016
2017	2013	2014	2015	2016	2017	2013	2014	2015	2016	2017
2018	2014	2015	2016	2017	2018	2014	2015	2016	2017	2018
2019	2015	2016	2017	2018	2019	2015	2016	2017	2018	2019
2020	2016	2017	2018	2019	2020	2016	2017	2018	2019	2020
2021	2017	2018	2019	2020	2021	2017	2018	2019	2020	2021
2022	2018	2019	2020	2021	2022	2018	2019	2020	2021	2022

For **Maternal Deaths**, due to the concern of small number of death events, maternal mortality was calculated for the whole 13-year period 2010-2022.

Censoring Values

Censoring criteria based on the data use agreement with the National Center for Health Statistics (NCHS). In accordance with our data use agreement with the NCHS, if the total count of events is less than 10, then the BCHI cannot report the value.

Internal censoring criteria. Some data were also suppressed based on BCHI platform criteria. BCHI platform criteria were developed after extensive review of the calculated rates, alignment with mortality trends reported by other health agencies, and following best practices for reporting data with small cases.

1) Prenatal Care. Prenatal care was censored for Phoenix, AZ (for years 2010 to 2014), Boston, MA (for years 2010 and 2011), Charlotte, NC (for years 2010 and 2011), Minneapolis, MN (for years 2010 and 2011) because prenatal care information were not included in the birth certificates for these years.
2) Preterm Births. Preterm Births was censored for all cities for years 2010 to 2014 due to the change in the way gestational age was calculated prior to 2014.
3) Opioid Overdose Deaths. Opioid overdose deaths for Philadelphia, PA were suppressed for year 2010 to 2017 because of reporting issues with overdose deaths in the death certificates for these years.
4) Mortality by Race Strata. Manual review further removed a small number of metric-city-race group values when the city race population was low (race population <40k) and the metric-city-race values showed highly erratic trends that zig-zagged far outside the bounds of the rest of the metric-city-race values. The assumption being that the erratic values are most likely artifacts of the data driven by the unstable population estimates and did not reflect actual fluctuation in trends. Others have used population thresholds to suppress values that are thought to be artifacts of the data. See Cathy Wasserman, Eric Ossiander,

Department of Health for Washington State (2018) "Standards for Reporting Data with Small Numbers" doh.wa.gov/sites/default/files/legacy/Documents/1500//SmallNumbers.pdf.

5) Mortality by Race and Sex Strata: The following automatic criteria were implemented to remove a small number of metric-city-race-sex group values for metrics with high prevalence in cause of death, namely: Deaths from All Causes, All Cancer Deaths, Lung Cancer Deaths, Cardio-vascular Disease Deaths, Heart Disease Deaths, Diabetes Deaths, Pneumonia and Influenza Deaths, and Colorectal Cancer Deaths. For those, metric-city-race-sex group values were censored if the rate deviated 1.5 standard deviation (STD) from the mean race-sex rate for all cities. The assumption being that common causes of death should have sufficient numbers to withstand sub-dividing by metric-city-race-sex, given that in the past 10 years, there has been more-or-less stable medical interventions and policy-related changes would have only gradual impacts given that the rates used 3-years of data.

Values for the other metrics-city-race-sex were left as-is (not censored using an internal criteria).

Mortality Rates

Multiple Causes of Mortality restricted-use micro-data files from the NVSS were used for calculation of mortality rates.

Causes of death were classified in accordance with World Health Organization (WHO) regulations based on current revision of the International Statistical Classification of Diseases and Related Health Problems (ICD 10)[6]. The table of ICD 10 codes for different causes of mortality are given in **Table 6**.

Table 6: List of ICD 10 codes used for Mortality.

Metrics	ICD-10 Codes
All Cancer Deaths	C00-C97
Breast Cancer Deaths (calculated for females)	C50
Cardiovascular Disease Deaths	I00-I78
Colorectal Cancer	C18-C21
Covid-19 Deaths	U07.1
Deaths from All Causes	All deaths included
Diabetes Deaths	E10-E14
Drug Overdose Deaths	F11.0-F11.5, F11.7-F11.9, F12.0-F12.5, F12.7-F12.9, F13.0-F13.5, F13.7-F13.9, F14.0-F14.5, F14.7-F14.9,
	F15.0-F15.5, F15.7-F15.9, F16.0-F16.5, F16.7-F16.9,
	F17.0, F17.3-F17.5, F17.7-F17.9, F18.0-F18.5,
	F18.7-F18.9, F19.0-F19.5, F19.7-F19.9, X40-X44, X60-X65, X85, Y10-Y14

Metrics	ICD-10 Codes
Firearm-Related Deaths	W32-W34, X72-X74, X93-X95, Y22-Y24, Y35.0
Heart Disease Deaths	I00-I09, I11, I13, I20-I51
HIV-Related Deaths	B20-B24
Homicides	X85-Y09, Y87.1
Injury Deaths	V01-X59, Y85-Y86 X60-X84, Y87.0 X85-Y09, Y87.1 Y35, Y89.0 Y10-Y34, Y87.2, Y89.9
Lung Cancer Deaths	C34
Maternal Deaths	A34, O00-O95, O98, O99
Motor Vehicle Deaths	V02-V04, V09.0, V09.2, V12-V14, V19.0-V19.2, V19.4-V19.6, V20-V79, V80.3-V80.5, V81.0-V81.1, V82.0-V82.1, V83-V86, V87.0-V87.8, V88.0-V88.8, V89.0, V89.2
Opioid Overdose Deaths	Opioid overdose deaths are identified by the presence of any of the following Multiple Cause of Death (MCOD) codes in addition to the Underlying Cause of Death (UCD) codes X40-X44, X60-X65, X85, Y10-Y14: T40.0(opium); T40.1(heroin); T40.2 (natural opioid analgesics); T40.3 (methadone); T40.4 (other synthetic); or T40.6 (other and unspecified narcotics).
Pneumonia or Influenza Deaths	J09-J18
Prostate Cancer	C61
Suicide	X60-X84 and Y87

Population: American Community Survey 1-year population estimates were used as population denominators for mortality rates. Age-group specific population by sex were obtained from tables B01001 (all races), B01001B (Black), B01001D&E (Asian/PI), B01001H (non-Hispanic White), and B01001I (Hispanic).

Infant Deaths and Maternal Deaths are reported per number of live births. Number of live births are calculated from the Natality restricted-use micro-data files from the NVSS.

Weights: The mortality rates (except for Infant Deaths and Maternal Deaths) are age-adjusted by the direct method of age adjustment using the US 2000 standard population. All deaths where age was missing, unknown, or not stated are excluded from the analysis.

Metric Calculation:

$$\text{Age-adjusted death rate} = \sum \frac{\text{Deaths from specific cause for age-group}}{\text{Total population for age-group}} * \text{age-group weight} * 100{,}000$$

$$\text{Infant Mortality Rate} = \frac{\text{Total infant deaths}}{\text{Total live births}} * 1000$$

$$\text{Maternal Mortality Rate} = \frac{\text{Total maternal deaths}}{\text{Total live births}} * 100{,}000$$

Birth Outcome

Natality restricted-use micro-data files from the NVSS were used for calculation of birth outcome. Birth outcome metrics are not age-adjusted. Birth weight, age of mother, gestational period, and prenatal care information was given in the birth records.

Population: 'Low Birth Weight', 'Preterm Births' and 'Prenatal Care' are reported per number of live births. Number of live births are also calculated from the Natality restricted-use micro-data files from the NVSS. 'Teen Birth' is reported per unit female population aged 15-19 years. Female population estimates were obtained from tables B01001 (all races), B01001B (Black), B01001D&E (Asian/PI), B01001H (non-Hispanic White), and B01001I (Hispanic) of the American Community Survey 1-year survey.

Metric Calculation:

$$\text{Low Birthweight} = \frac{\text{All births-weights <2500 grams}}{\text{Total live births}} * 100$$

$$\text{Prenatal Care} = \frac{\text{All births where prenatal care began in first 3 months}}{\text{Total live births}} * 100$$

$$\text{Preterm Birth} = \frac{\text{All births before 37 completed weeks of gestation}}{\text{Total live births}} * 100$$

$$\text{Teen Births} = \frac{\text{Births where age of mother is 15-19}}{\text{Female population 15-19 years}} * 1{,}000$$

HEAT WAVE METRICS FROM THE GLOBAL HISTORICAL CLIMATOLOGY NETWORK DAILY DATABASE

General Notes:

Temperature data were sourced from weather station data compiled by the National Oceanic and Atmospheric Administration (NOAA) Global Historical Climatology Network daily (GHCNd) database.

Our heat wave metric calculations followed definitions and methodology used by U.S. Environmental Protection Agency National Centers for Environmental Information (EPA-NCEI), which they adapted from Habeeb et al. (2015) [7].

Choosing weather stations:

We selected a single weather station for each city using the following inclusion criteria:

1) Station data completeness met those established by Habeeb et al. (2015):

a) Station had temperature data for each year of the observation period (in our case, 1960 to 2022).

b) No more than 1.6% of months were missing (in other words 98.4% were not missing). For our observation period 1960 to 2022, this meant that only 12 months could be missing. A station-month was considered missing if there were no temperature data for more than 15 days in the month.

2) Stations were able to approximate temperature values in the city. Nearly all selected stations were 2 miles to 10 miles from the center of the city.

Heat Wave Definition:

There is no universally accepted definition for a heat wave. We used EPA-NCEI's heat wave definition (defined in italics below) except that our calculation used daily minimum dry bulb temperature instead of daily minimum apparent temperature. We used dry bulb temperature because relative humidity -- which is used to derive apparent temperature -- was unavailable for years prior to the 2000's (and a few city weather stations lacked relative humidity even in the 2000's). City heat wave counts may not be highly sensitive to whether minimum dry bulb or minimum apparent temperature is used because heat wave is defined relative to typical weather conditions the city (as described below).

EPA-NCEI's heat wave definition and rationale are quoted below from their documentation.

They defined heat wave as "*a period of two or more consecutive days where the daily minimum (apparent) temperature at a particular weather station is higher than the 85th percentile of historical July and August temperatures for that city. Historical July and August baseline temperatures are analyzed for a base period of 1981–2010, which NCEI chose for consistency with other climatology metrics. Using the 85th percentile of July and August temperatures results in a threshold that equates to the nine hottest July and August days in an average year during that 30-year window. These are likely among the nine hottest days of the*

year. A temperature that is typically only recorded nine times during the hottest two months of the year is rare enough that most people would consider it to be unusually hot. Using city-specific thresholds rather than a single nationwide threshold (e.g., 95°F) ensures that this indicator accounts for local variations in conditions, and it acknowledges differences in the extent to which people have acclimated or adapted to high temperatures. For example, a 95°F Day in Milwaukee could arguably pose a more severe health risk than a 100°F day in Phoenix, given the greater prevalence of air conditioning in Phoenix and the extent to which Phoenix residents are accustomed to hot weather. "

Heat Wave Season Length was defined using definitions applied by EPA-NCEI: the number of days from the *first day* of the first heat wave to the *last day* of the last heat wave.

Details on metrics

Number of Heat Waves was calculated as the average number of heat waves per year during the period 2010-2022. Years 2010-2022 were chosen because these are the years that are covered by the BCHC data platform. Similar to EPA-NCEI's methods, we averaged the number of heat waves across multiple years because weather conditions (and number of heat waves) vary widely year-to-year.

Metric calculation:

$$\text{Average Number of Heat Waves 2010 to 2022} = \frac{\text{Total number of HW for 13 years from 2010 to 2022}}{\text{13 years}}$$

Longer Summer (AKA Longer Heat Wave Season) was defined as an increase in the heat wave season length during the past 60 years. EPA-NCEI used this metric when describing trends in heat waves across the United States. The major difference between our calculation and EPA-NCEI's is that our calculations simply derived the difference between the average heat wave season length in the latest period to the average season length in the earliest period, whereas EPA-NCEI's used a linear prediction of the trend (change) across decades.[1]

Metric calculation:

Longer Heat Wave Season =

Average Heat Wave Season Length in later period (2010 to 2022) minus Average Heat Wave Season Length in early period (1960 to 1969)

[1] EPA's decision to use a model was likely motivated by having very few years in the most recent period when they performed their analyses. The model trend estimate less interpretable than simple differencing of the earliest vs most recent season length.

DISASTER DECLARATION DATABASE OF THE FEDERAL EMERGENCY MANAGEMENT AGENCY (FEMA)

Disasters were identified via the Federal Emergency Management Agency's (FEMA), Disaster Declaration Database (National Emergency Management Information System). We retained disaster types that were related to climate: severe storms, hurricanes, floods, coastal storms, tornadoes, severe ice storms, freezing, snowstorms, droughts, and wildfire. Heat waves were not included because they are not currently a category that jurisdictions can request for federal declarations.

Disasters are declared in situations of catastrophic events that overwhelm the resources of local and state authorities. The governor of the state in which the disaster occurs declares a state of emergency, then formally requests assistance from the President of the U.S. and FEMA, then the federal government decides regarding the course of action and responds to the disaster.

We used federally declared disasters that were declared for the state or county either because (a) a catastrophic event occurred within the county (this is the most common reason for a declaration); or (b) a catastrophic event occurred outside of the county but within the county's disaster response region and the county needed to provide a high level of aid to the disaster response (Austin TX -- which is the state capital --is an example of a jurisdiction that fulfills this role).

Details on metric

We calculated the average number of days per year during the past two decades (2001 to 2022) that each city was affected by a federally declared climate-related disaster. Similar to other weather-related metrics, it was necessary to average disasters across many yrs because catastrophic climate events infrequently occurred in some BCHC jurisdictions.

Metric calculation:

Average number of days per year that city was affected by a federally declared climate-related disaster (2001 to 2022) =

$$\frac{\text{Total number of days under a disaster declaration for 22 years (2001 to 2022)}}{\text{22 years}}$$

Data cleaning. Major disasters can be declared at various stages of the request process, and during reassessment of the damage; thus, a single disaster event can have multiple entries in the FEMA database along with slightly different start- and end-dates. We de-duplicated the database in order to create a list of unique disasters. When a single event was represented by multiple entries with varying dates, we determined the disaster duration using the earliest start-date and the last end-date. New York City is represented by 5 counties (boroughs). We combined disaster information for the 5 New York City counties and then deleted duplicate disaster occurrences.

Appendix 3. Data download code book

This is the code book for the data download.

Field	Label
metric_item_label	Metric title
metric_cat_label	Metric category title
metric_subcat_label	Metric subcategory title
metric_item_label_subtitle	Metric subtitle; provides information about the metric.
metric_cat_item_yaxis_label	Label for y-axis units
metric_source_desc_label_fn	The source of data.
metric_source_desc_label_url_fn	URL for the source of data.
geo_label_city	City name. Also contains "U.S. Total".
geo_label_state	State abbreviation
geo_label_citystate	City name, state abbreviation. Also contains "U.S. Total"
geo_fips_code	5 (County FIPS code) or 7 (Census Places FIPS code) digit unique number for the location
value	The data value for the metric
date_label	The label for 'year' that is displayed in the platform visualizations
geo_label_proxy_or_real	This field indicates whether the source data geographic unit is 'city' level or something else. "real" = city; "proxy" = means it is county, MSA, a city's school district, etc
geo_label_proxy_footnote	If "proxy" geographic unit is used, the geographic unit is noted here (eg. county, MSA etc.)
geo_fips_desc	Type of geographic unit being represented: place, county, country (Note "Place" is the census term for "city")
date_label_proxy_or_real	This field indicates whether the source data represent the year labeled In "date_label" or whether the year is proxied via interpolation or imputed from another year. "real" = data is for the date_label year; "proxy" = data are imputed from another year.
date_label_proxy_footnote	Footnote text that is displayed if proxy data is used for that year.
value_ci_flag_yesno	Flag whether confidence interval is present
value_95_ci_low	95% confidence interval, lower bounds
value_95_ci_high	95% confidence interval, upper bounds
value_90_ci_low	90% confidence interval, lower bounds
value_90_ci_high	90% confidence interval, upper bounds
geo_strata_region	Census region the city belongs to
geo_strata_poverty	Poverty strata the city belongs to
geo_strata_Population	Population strata the city belongs to
geo_strata_PopDensity	Population density strata the city belongs to
geo_strata_Segregation	Racial Segregation strata the city belongs to
strata_race_label	Race/ethnicity values: 'Black', 'White', 'Hispanic', 'Asian/PI', 'American Indian/Alaska Native' etc.
strata_sex_label	Sex values: Male, Female
strata_race_sex_label	Race/ethnicity & Sex values: 'Black Male', 'Black Female', 'White Male' etc.

Appendix 4. Citing bigcitieshealth.org and references used in the technical documentation

HOW TO CITE BIGCITIESHEALTH.ORG

We encourage use of platform data and visualizations, and suggest the following citation:

> Big Cities Health Coalition and Drexel University Urban Health Collaborative. 2024 Data Platform and Technical Documentation: Big Cities Health Inventory Available at www.bigcitieshealth.org . Accessed [INSERT DATE OF ACCESS]

If you have questions or suggestions regarding the data platform, please submit them on this web form.

REFERENCES FOR THE TECHNICAL DOCUMATION

1. Missouri Census Data Center. (n.d.). Retrieved September 29, 2021, from https://mcdc.missouri.edu/applications/geocorr.html.
2. Dranger E, Remington P. Years of Potential Life Lost (YPLL): A summary measure of premature mortality used in measuring the health of communities 2004.
3. Sprague, Thomas B. "Explanation of a new formula for interpolation." Journal of the Institute of Actuaries 22.4 (1880): 270-285.
4. National Center for Health Statistics. All Counties Natality files, 2010-2021, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program.
5. National Center for Health Statistics. Detailed Mortality- All Counties files, 2010-2022 as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program.
6. National Center for Health Statistics. Compressed Mortality File, 1999-2011 (machine readable data file and documentation, CD-ROM Series 20, No. 2Q) as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program. Hyattsville, Maryland. 2014.
7. Habeeb, D., Vargo, J., & Stone, B. (2015). Rising heat wave trends in large US cities. Natural Hazards, 76, 1651-1665.