Data repository: Technical Assumptions

## General rules/assumptions:

#1- **Data Aggregated from County Level to Larger Geography**: Certain data points were aggregated from county level to another larger geography for example MSA level or state level.

#2- **Aggregation with Missing Data Points**: If any county data was missing or unavailable, data was still aggregated as long as 80% of the MSA’s population could be accounted for with available data. This was done using the variables labeled “completeness.”

#3- **Restratify Age Counts**: Re-calculated given age groupings from a dataset to meet the age groupings used in the SHIELD model.

#4- **Redistributed Racial Data**: Re-distributed certain racial groups among other racial groups to meet the race groupings for SHIELD. This was done for multiracial, unknown, or more than one race.

#5- **Race Recalculation**: Applied this calculation to parse out certain racial groups when the data was unavailable. Used available proportions of racial data to create estimates. Calculated as:

Black, Non-Hispanic = Total Black – (sqrt(proportion.black.hispanic \* proportion.hispanic.black \* Hispanic.value \* black.value)) and other.race = (total – (Hispanic.value +black.nh)

# Outcome category: Population data

Output: Population

* Description: Count of individuals.
* Source: [1]
* Assumptions: #1

Output: Deaths

* Description: Count of general mortality.
* Source:[1]; [2]
* Assumptions: #1, #4

Output: Immigration

* Description: A count of individuals who moved from abroad to the United States in given year.
* Source: [3], [4]
* Assumptions: #3, #5

Output: Emigration

* Description: A count of individuals who moved out of an MSA in a given year.
* Source: [3]
* Assumptions: #3, #5

# Outcome category: Fertility data

* 1. Output: Fertility Rate
* Description: Number of births per 1,000 women of childbearing age (15-44).
* Source: [5]
* Assumptions: Aggregated to MSA level by first aggregating the numerator (births numerator for fertility rate) and denominator (female population denominator for fertility rate) and then dividing to get the fertility rate calculation. Calculated as (births/female.population aged 15-44). Removed ‘unidentified counties’ (counties with a population less than 100,000). Combined Asian, Pacific Islander, and Native Hawaiian into one racial group so that CDC’s Bridged Race data could be used along with Single Race data.
  1. Output: Births Numerator for Fertility Rate
* Description: Count of live born infants.
* Source: [5]
* Assumptions: #1, #2. Removed ‘unidentified counties’ (counties with a population less than 100,000). Combined Asian, Pacific Islander, and Native Hawaiian into one racial group so that CDC’s Bridged Race data could be used along with Single Race data.
  1. Output: Female Population Denominator for Fertility Rate
* Description: Count of women in fertile window, ages 15-44.
* Source: [5]
* Assumptions: #1, #2. Removed ‘unidentified counties’ (counties with a population less than 100,000). Combined Asian, Pacific Islander, and Native Hawaiian into one racial group so that CDC’s Bridged Race data could be used along with Single Race data. Used the ‘Completeness Female Population Denominator for Fertility Rate’ variable to assess that at least 80% of the population of the MSA was represented by the available data.
  1. Output: Completeness Female Population Denominator for Fertility Rate
* Description: A variable used to assess if there is adequate data available to justify aggregating female population data from county level to MSA level.
* Source: [5]
* Assumptions: Calculation is (Total Summed County Population for Counties that have Female Population Data/Total Population of the MSA)
  1. Output: No Prenatal Care; Prenatal Care Initiation First Trimester; Prenatal Care Initiation Second Trimester; Prenatal Care Initiation Third Trimester
* Description: Proportion of women ages 15-44 who initiated prenatal care in various trimesters of their pregnancy (first, second, third, or not at all).
* Source: [6]
* Assumptions: #1, #2. Removed unknown or not reported data points. Only included women ages 15-44. Used the ‘Completeness Prenatal Care Initiation First Trimester’ variable to assess that at least 80% of the population of the MSA was represented by the available data. Calculated as (count of women who initiated care in a given trimester / births that year).
  1. Output: Completeness Prenatal Care Initiation First Trimester
* Description: A variable used to assess if there is adequate data available to justify aggregating prenatal care data from county level to MSA level.
* Source: [6]
* Assumptions: Calculation is (Total Summed County Population for Counties that have Prenatal Care Data/Total Population of the MSA)

# Outcome category: HIV Data

* 1. Output: HIV Diagnoses
* Description: HIV Diagnoses refer to the number of HIV infections confirmed by laboratory or clinical evidence in the calendar year.
* Source: [7]
* Assumptions: #1
  1. Output: HIV Diagnosed Prevalence
* Description: This is a count of individuals who are living with HIV. This prevalence value is based only off of people who have HIV who know that they have HIV.
* Source: [7]
* Assumptions: #1
  1. Output: HIV Total Prevalence
* Description: This value is an estimation of every individual who has HIV (whether they are aware of their status or not).
* Source: [7]
* Assumptions: none
  1. Output: PrEP
* Description: Number of people prescribed PrEP in a calendar year.
* Source: [7], [8]
* Assumptions: #1, #4
  1. Output: PrEP Indications
* Description: Estimated number of individuals with indications for PrEP use.
* Source: [7]
* Assumptions: #1, #4
  1. Output: HIV Suppression
* Description: The proportion of individuals living with HIV who are virally suppressed. Only includes individuals alive at the end of the queried year.
* Source: [7]
* Assumptions: #1
  1. Output: HIV Engagement
* Description: Proportion of persons age 13+ who received an HIV diagnosis by the end of the previous year who resided in a jurisdiction with complete reporting of CD4 and viral load results to CDC. Commonly referred to as ‘receipt of care.’
* [7]
* Assumptions:
  1. Output: Proportion MSM
* Description: An estimated proportion of individuals who are men that have sex with men.
* Source: EMORY, [9]
* Assumptions: Calculated variable. It encompasses anyone who reported to BRFSS that their sex is ‘male’ and their sexual orientation is ‘gay’ or ‘bisexual.’ Applied BRFSS weighting (\_LLCPWT) when processing BRFSS data.
  1. Output: Proportion Tested for HIV; Proportion Tested for HIV n
* Description: An estimate of the proportion of individuals who were tested for HIV in the past year. Proportion Tested for HIV N is the total count of individuals (denominator value for the proportion).
* Source: [9]
* Assumptions: To calculate these variables, observations were removed if date of last HIV test was refused. BRFSS only reports month and year of past test. If test month was missing but year was available, it was assumed the test was done in June as this assigns a 50/50 probability that the test was truly in the past year.

# Outcome category: Syphilis Data

* 1. Output: PS Syphilis Diagnoses
* Description: Diagnoses of syphilis in the primary or secondary stage. These are the surveillance stages that represent the symptomatic and earliest stages of infection.
* Source: [7], [10], Local Health Department Reports
* Assumptions: #1
  1. Output: Primary Syphilis Diagnoses
* Description: Count of diagnoses during the stage of syphilis infection clinically characterized by a chancre.
* Source: [11]
* Assumptions: none
  1. Output: Secondary Syphilis Diagnoses
* Description: Count of diagnoses during the stage of syphilis infection clinically characterized by a rash.
* Source: [11]
* Assumptions: none
  1. Output: Early Syphilis Diagnoses
* Description: Cases of syphilis that likely occurred in the prior 12 months but do not have signs and symptoms of primary or secondary syphilis.
* Source:[7], [10], Local Health Department Reports
* Assumptions: #1
  1. Output: Total Syphilis Diagnoses
* Description: A total count of syphilis diagnoses that includes the summation of primary, secondary, early, unknown duration, and congenital cases. Also includes stage of syphilis not stated.
* Source: [10], [11], Local Health Department Reports
* Assumptions: none
  1. Output: Congenital Syphilis Diagnoses
* Description: Count of babies born with syphilis as passed from mother infected with syphilis.
* Source: [7], [10], Local Health Department Reports
* Assumptions: none
  1. Output: CNS Syphilis Diagnoses
* Description: Neurosyphilis- count of syphilis cases where there is evidence of syphilis in the central nervous system.
* Source: [12]
* Assumptions: none
  1. Output: Unknown Duration or Late Syphilis Diagnoses
* Description: Count of syphilis cases that have occurred more than 12 months prior or cases where there is not sufficient evidence to determine that the case was acquired in the past 12 months.
* Source: [7], [10], [12], Local Health Department Reports
* Assumptions: #1
  1. Output: Proportion of Congenital Syphilis Births
* Description: The proportion of total births that are cases of congenital syphilis. This is calculated as (# of congenital syphilis births/ # total births)
* Source: [7], Local Health Department Reports
* Assumptions: Denominator is ‘births.denominator.for.congenital.syphilis.proportion’ where data is from CDC Wonder.
  1. Output: Births Denominator for Congenital Syphilis Proportion
* Description: Count of liveborn infants.
* Source: [5], [13], [14]
* Assumptions: #1. To calculate the ‘proportion.of.congenital.syphilis.births,’ birth were aggregated from county to MSA (#1) prior to calculating the ‘proportion.of.congenital.syphilis.births.’

1. Bureau USC. Population and Housing Unit Estimates Datasets.

2. Centers for Disease Control and Prevention NCfHS. National Vital Statistics System, Mortality 1999-2020 on CDC WONDER Online Database, released in 2021.

3. (ACS) USCBaACS. Metro-to-Metro Migration Flows.

4. Survey USCBaAC. Geographical Mobility in the Past Year by Sex for Current Residence in the United States.

5. Centers for Disease Control and Prevention NCfHS. National Vital Statistics System, Natality on CDC WONDER Online Database. Data are from the Natality Records 2007-2023, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program.

6. Centers for Disease Control and Prevention NCfHS. National Vital Statistics System, Natality on CDC WONDER Online Database. Data are from the Natality Records 2016-2023, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program.

7. Prevention CfDCa. NCHHSTP AtlasPlus.

8. AIDSVu: Sullivan PS WC, Koski C,, Pembleton E MP, Taussig J, Ricca A, Luisi N, Mokotoff E, Benbow N,, Castel AD DA, Valdiserri RO, Bradley H, Jaggi C, O’Farrell D, Filipowicz R,, Siegler AJ CJ, Sanchez TH. A Data Visualization and Dissemination Resource to

Support HIV Prevention and Care at the Local Level: Analysis and Uses of the

AIDSVu Public Data Resource. J Med Internet Red, **2020**.

9. *Behavioral Risk Factor Surveillance System Survey Data* . In: (CDC) CfDCaP. Atlanta, Georgia: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

10. Prevention CfDCa. Sexually Transmitted Infections Surveillance 2022. Atlanta: US

Department of Health and Human Services, **2024**.

11. Prevention CfDCa. *Sexually Transmitted Disease Surveillance 2017*. Atlanta: U.S. Department of Health and Human Services, **2018**.

12. Disease Cf, Prevention Ca. *Sexually Transmitted Disease Surveillance, 2003*. Atlanta: U.S. Department of

Health and Human Services, **2004**.

13. Centers for Disease Control and Prevention, Statistics NCfH. National Vital Statistics

System, Natality on CDC WONDER Online Database.  . **2003-2006**.

14. Centers for Disease Control and Prevention, Statistics NCfH. National Vital Statistics System, Natality on

CDC WONDER Online Database. **1997-2002**.