RESOURCE SHEET:

USING CENSUS AND ACS DATASETS

U.S. Census Bureau data products that include important measures of the social, economic, demographic, and housing characteristics of the U.S. population are: 1) Decennial Census and 2) American Community Survey (ACS).

TOPICS COVERED:

WHAT IS THE DIFFERENCE BETWEEN DECENNIAL CENSUS AND ACS?

The decennial census is conducted every 10 years in the years ending with "0" (i.e. 2000, 2010, 2020). The decennial census is used to count every person living in the United States and obtain basic information about age, sex, race, Hispanic origin, and owner/renter status. Prior to 2010, the decennial census had an additional survey administered to a sub-sample that contained additional questions pertaining to social and economic conditions such as education, employment, and income to name a few. This additional survey is no longer included in the decennial census starting in 2010.

The American Community Survey (ACS) is an ongoing survey conducted every month in every year starting in 2005 through present time. This survey is sent to a sample of addresses across the U.S. and contains questions about social, economic, demographic, and housing characteristics. This survey replaced the survey that was given to the additional sub-sample in the decennial census prior to 2010. Questions that were asked of the sub-sample are no longer available in the decennial census data after 2010 but are now available in the ACS data.

Although the ACS is administered every year, the data available uses multiple years aggregated together to obtain reliable estimates for small areas due to the smaller sample sizes. For geographies of less than 20,000 persons (census tract and block group), the data are only available as 5-year aggregated data (for example: tract data are available for 2005-2009 but not 2009 alone). This means that the estimates represent the entire 5-year period. For geographies with more than 20,000 persons, including many counties, ACS data are available for 1-year estimates and are available for every year from 2005-2021. UHC has currently processed the ACS 5-year aggregate estimates. The official date associated with the decennial census data is April 1 of the year when the census was administered (i.e: 2010 census asks respondents where

they lived on April 1, 2010). More information about the ACS 1-year and 5-year aggregated data can be found at: When to Use 1-year or 5-year Estimates and Period Estimates in the American Community Survey.

WHAT MEASURES ARE AVAILABLE?

The census and ACS data cover a broad range of topics about social, economic, demographic, and housing characteristics.

When data are obtained from the census bureau, it is generally available as counts that are then used to create percentages. For example, if you are interested in poverty, the census bureau's data will have the number of people below the poverty line and the total population who answered the poverty-related questions. To use these data for analyses or reporting, you should report the percentage living in poverty, not total numbers, to give better comparisons across areas.

Census and ACS data are associated with a "universe" – the population included for a particular count. Examples of universes are: total population, population age 25 and above (education attainment), families, households, and housing units.

UHC has processed census and ACS datasets and created many measures that are of general interest in urban health research. Measures created are:

- Land area, total population, total number of households, total number of families, population density
- Population counts by age-sex-race/ethnicity
- Age distribution (% less than 5 years; % less than 18 years; % 65 and older)
- Sex distribution (% male)
- Race/Ethnicity (% Hispanic, % white non-Hispanic, % Black non-Hispanic, % Asian non-Hispanic, % other races non-Hispanic, % multi-racial non-Hispanic)
- Place of birth (% foreign born, % Puerto Rico, % Africa, % north Africa, % Latin America, % China)
- Single parent households (% female only with children, % male or female only with children)
- Housing Stability (% who moved into current house by time period, % lived in same county in past year)
- Education-overall and by race/ethnicity (% with highest level of education high school degree, % with bachelor's degree or higher, % with some college/associates degree)
- School enrollment (% nursery or preschool, % college, % 16-19 not enrolled in school)
- Language spoken at home (% speak Spanish, % speak Chinese)
- Poverty-overall and by race/ethnicity (% below poverty level for all ages, ages 5 and older, age<18 years, ages 18-64, ages 65+)
- Income to poverty ratio
- Income (median household income, % with investments, % with social security, % with public assistance)
- Income per capita-overall and by race/ethnicity
- Employment-overall and by sex, race/ethnicity (% employed, % unemployed, % not in labor force)
- Occupation (% management, % service occupations, % sales/office, % farming/fishing/forestry, % construction, % production/transportation)
- Commuting (% travel time to work <15 minutes, <30 minutes, >=60 minute; % transport type of drove alone, carpooled, public transportation, biking, walking)
- Health Insurance-for all ages and children (% public, % private, % uninsured)
- Home occupied and ownership (% occupied, % vacant, % owner occupied, renter-to-owner occupied ratio, average household size, % with overcrowding)

- Housing cost (median rent, median rent as % of income, median home value, % spending >=35% of income on rent/mortgage)
- Housing conditions (% without telephone, % without vehicle, % without plumbing)

Although UHC has a comprehensive list of measures, there may be other topics of interest. More information about data collected can be found at: Why We Ask Each Question and Explore Census Data.

WHAT GEOGRAPHIC LEVELS ARE AVAILABLE?

UHC has created measures for the census geographies of places, county, census tract, census block group, and zip code tabulation areas.

Counties are primary political and administrative divisions of states. Census tracts (CT) are small statistical subdivisions of counties. These are designed to provide a set of geographic units for decennial census data. CT typically have populations of 1,500-8,000 persons with an optimal size of 4,000. CT are often used as proxies for "neighborhoods" or "small areas" for urban health research. Block groups (BG) are statistical divisions of census tracts. BG typically have populations of 600-3,000 persons with an optimal size of 1,500. Zip code tabulation areas (ZCTA) are approximate areas of U.S. Postal Service zip codes created from census blocks. Zip codes are not a census designated geography, so they do not have census data available. If your research uses zip codes, it is recommended that you use ZCTA as an approximation. Census blocks are the smallest geographic units used by the census and the building blocks for all census geographic boundaries. Decennial census counts are available at the census block level, but ACS data are not available at the block level.

Other geographic levels, such as states, tribal areas, school districts, metropolitan areas, are available in the census and ACS data but have not been processed by UHC.

More information about census geographies can be found at: Census Geographies

THERE IS MISSING DATA. WHAT DOES THAT MEAN?

Although the decennial census and ACS sample the entire U.S., missing data are found for some of the estimates and data provided. The Census Bureau suppresses data to limit disclosure of information when it could identify individuals, such as when counts within certain subgroups are small, or estimates from the survey samples have unacceptable levels of statistical reliability. Some census tracts are designated special land use tracts which may be missing data. For example, tracts that have all residents in group quarters such as prisons will not have household or family data.

More information about data suppression can be found at Data Suppression.

MARGINS OF ERROR IN ACS ESTIMATES

ACS estimates are subject to sampling variability. The US Census reports estimates with their corresponding margins of error, allowing for the calculation of 90 percent confidence intervals. The UHC does not include margins of error on the portal. These can be obtained directly from the US Census.

More information margins of error can be found at Guidance for Data Users American Community Survey (ACS).

CAN I COMPARE CENSUS AND ACS DATA ACROSS TIME?

Yes, but there are some considerations for comparing across time. When using the ACS 5-year aggregate estimates, the Census Bureau recommends only comparing 5-year periods that are not overlapping because much of the data are the same sample. For example, you can compare 2006-2010 to 2011-2015 but you should not compare 2006-2010 to 2007-2011. A good example explaining why you should only compare non-overlapping periods is found at Period Estimates in teh American Community Survey

Every 10 years, the Census Bureau redraws the boundaries of census tracts and block groups. Occasionally corrections are made to the decennial census boundaries after census years and IDs are changed to accommodate administrative boundary changes (e.g. changes in county boundaries). Changes in administrative boundaries by agencies other than the census can occur at any time, such as changes in incorporated places such as cities, and changes in metropolitan statistical areas (core based statistical areas, combined statistical areas). Changes and decennial census boundaries and other administrative boundary changes can be found at TIGER/Line Geodatabases.

Occasionally the census changes the wording of questions. Most notably the census changed the way questions about race and ethnicity were asked in 2000, including allowing people to answer two or more races. The Census Bureau outlines some improvements to race questions here: Improvements to the Race Question

A comparison of the race/ethnicity questions over time is provided by IPUMS USA:

- Race: https://usa.ipums.org/usa-action/variables/RACE#description_section
- Hispanic ethnicity: https://usa.ipums.org/usa-action/variables/HISPAN#description_section

IPUMS USA is a good resource for comparing most census/ACS questions over time.

More information about comparing across time can be found at: Comparing ACS Data

I HEARD THERE ARE PROBLEMS WITH THE 2020 CENSUS AND ACS. IS THAT TRUE?

The Census Bureau has acknowledged that there are likely under- and overcounts of certain populations in the 2020 Census data. Specifically, they believe that there is an undercount of Black/African American, Native American and Native Alaskan living on a reservation, Hispanic/Latino population, and those reporting some other race and non-Hispanic White and Asian populations are overcounted. They also believe that children 0-17 years old, particularly those under age 5, are undercounted.

The 2020 ACS data collection was disrupted by the COVID-19 pandemic. Due to the data collection restraints, the Census Bureau's standard weighting methodology was found to be inadequate for creating the ACS estimates. The ACS 2020 1-year estimates are released as an experimental product. Because the 5-year aggregate estimates starting in 2016-2020 use the responses from the experimental ACS 2020 data, the 5-year aggregate data should be used with caution at this time.

More information about the 2020 Census data can be found at: Census Bureau Releases Estimates of Undercount and Overcount in the 2020 Census

A report about the impact of the COVID-19 pandemic on the ACS 2020 data can be found at: An Assessment of the COVID-19 Pandemic's Impact on teh 2010 ACS 1-Year Data

WHAT DATA IS ON THE UHC DATA PORTAL?

The UHC Data Repository has census data from 2000 and 2010 decennial census and all ACS data starting in 2005-2009 to 2016-2020 for county, census tract, block group, and zip code tabulation area (ZCTA). Not all of the data we have in the UHC Data Repository are currently on the UHC Data Portal.

Currently, the UHC Data Portal contains two different versions of census and ACS files which are a subset of the UHC data repository.

1) The calculated social, economic, demographic, and housing characteristics measures (i.e. percentages) for census 2000 and 5-year estimates for 2005-2009, 2010-2014, 2015-2019 and 2016-2020. Most users of the census/ACS data will use these datasets, such as for social exposures for research, or reporting for community groups. These datasets can be found under the title "American Community Survey 5-year (20xx-20xx)"
2) Population counts overall and by age group-sex-race/ethnicity for census 2000, census 2010, and all ACS starting in 2005 through 2019 for census tracts (note that the race/ethnicity categories are slightly different for census data vs ACS data. See the variable lists for details.). These data can be used in cases where knowing the number of people is important such as denominators for mortality rates. These datasets can be found under the title "American Community Survey Population Estimates 5-year (20xx-20xx)" and "Decennial Census Population Estimates (20xx)"

Users have access to the SAS and CSV format for each dataset. The measures (variables) included in the datasets on the UHC Data Portal are found in the "variable lists" in CSV format.

WHO DO I CONTACT IF I HAVE QUESTIONS?

If you have any questions about the available data or would like to discuss your needs, please contact the Urban Heath Collaborative data team at uhcdata@drexel.edu.