12. GLOSSARY

Accuracy. One of four key dimensions of survey quality. Accuracy refers to the difference between the survey estimate and the true (unknown) value. Attributes are measured in terms of sources of error (for example, coverage, sampling, nonresponse, measurement, and processing).

Allocation. A commonly used approach to imputation (a statistical procedure to fill in missing responses) is known as hot-deck allocation, which uses a statistical method to supply responses for missing or inconsistent data from responding housing units or people in the sample who are similar. Certain values, such as a person's educational attainment, are more accurate when provided from another housing unit or from a person with similar characteristics. Allocation rates measure the proportion of values that required hot-deck allocation and are an important measure of data quality.

American Community Survey (ACS). The ACS is a nationwide survey designed to provide communities a fresh look at how they are changing. The ACS replaced the decennial census long form in 2010 and thereafter by collecting long-form type information throughout the decade rather than only once every 10 years. Full national implementation of the ACS began in 2005. Questionnaires are mailed to a sample of addresses to obtain information about household residents and the housing unit itself.

The U.S. Census Bureau produces social, economic, housing, and demographic estimates from the ACS in the form of 1-year and 5-year estimates based on population thresholds. The strength of the ACS is in estimating population and housing characteristics. It produces estimates for small areas, including census tracts and block groups and population subgroups.

Although the ACS provides population and housing unit estimates, the Census Bureau's Population Estimates Program produces and disseminates the official estimates of the population for the nation, states, counties, cities and towns, and estimates of housing units for states and counties. For 2010 and other decennial census years, the decennial census provides the official counts of population and housing units.

Application Programming Interface (API). API is a set of programs that allows an application to interact with other applications. The Census Bureau has developed the Census API, enabling developers to design Web and mobile apps to provide quick and easy access from sets of Census Bureau statistics.

Assignment. A type of imputation (a statistical procedure to fill in missing responses), assignment involves looking at other data, as reported by the respondent, to fill in missing responses. For example, when determining sex, if a person reports giving birth to children in the past 12 months, the Census Bureau verifies that the person is female. This approach also uses data as reported by other people in the household to fill in a blank or inconsistent field. For example, if the reference person is a U.S. citizen, a biological child with a blank response to citizenship is also assumed to be a citizen.

Block Group. A statistical subdivision of a census tract, generally defined to contain between 600 and 3,000 people and between 240 and 1,200 housing units, and the smallest geographic unit for which the Census Bureau tabulates sample data. A subdivision of a census tract (or, before 2000, a block numbering area), a block group is a cluster of blocks having the same first digit of their four-digit identifying number within a census tract.

Census Geography. A collective term referring to the types of geographic areas used by the Census Bureau in its data collection and tabulation operations. This Web page shows the geographic entities for which data are available from the ACS, which provides information down to the block group level.[79]

With connecting lines, the diagram in the "Geographic Hierarchy" section shows the hierarchical relationships between geographic types. For example, a line extends from states to counties because a state is comprised of many counties, and a county can never cross a state boundary.

If no line joins two geographic types, then an absolute and predictable relationship does not exist between them. For example, many places do not cross a county boundary (that is, only one county). However, some places extend over more than one county like New York City. Therefore, an absolute hierarchical relationship does not exist between counties and places, and any tabulation involving both of these geographic types may represent only a part of one county or one place.

Census Tract. A small, relatively permanent statistical subdivision of a county delineated by a local committee of census data users for presenting data. Census tracts nest within counties and their boundaries normally follow visible features, but may follow legal geography boundaries and other nonvisible features

[79] U.S. Census Bureau, American Community Survey (ACS), Concept & Definitions <www.census.gov/programs-surveys/acs/geography-acs/concepts-definitions.html>.

in some instances. Census tracts ideally contain about 4,000 people and 1,600 housing units.

Coefficient of Variation (CV). The ratio of the standard error (square root of the variance) to the value being estimated, usually expressed in terms of a percentage (also known as the relative standard deviation). The lower the CV, the higher the relative reliability of the estimate.

Comparison Profiles (CP). The Comparison Profiles show ACS data side-by-side from different data releases, indicating where there is a statistically significant difference between estimates. Comparison Profiles are currently available for 1-year estimates and 5-year estimates starting with the release of the 2011–2015 ACS 5-year data. The 5-year Comparison Profiles will compare data between two nonoverlapping 5-year periods. ACS Comparison Profiles begin with the letters "CP."

Confidence Interval. The sample estimate and its margin of error permit the construction of a confidence interval that represents the degree of uncertainty about the estimate. A 90 percent confidence interval can be constructed by adding and subtracting the published margin of error from the ACS estimate. A 90 percent confidence interval can be interpreted roughly as providing 90 percent certainty that the true number falls between the upper and lower bounds.

Confidentiality. The guarantee made by law (Title 13, U.S. Code) to individuals who provide census information regarding nondisclosure of that information to others. By law, the Census Bureau cannot disclose any private information that identifies a person or a business. Under federal law, the penalty for unlawful disclosure is a federal prison sentence of up to 5 years, a fine of up to $250,000, or both.

Consumer Price Index (CPI). The CPI program of the Bureau of Labor Statistics produces monthly data on changes in the prices paid by urban consumers for a representative basket of goods and services.

Controls. During the ACS weighting process, the official county-level population and housing unit estimates are used as controls. Weights are adjusted so that ACS estimates conform to these controls (but do not necessarily exactly match). This is done to improve person and housing unit coverage and to reduce the variability of the ACS estimates.

Certain published estimates, such as total population estimates for states, do exactly match the controls. These estimates, which have five asterisks (*****) in the Margin of Error column in data.census.gov, are by definition fixed, and can be considered to have no sampling error.

Current Residence. The ACS uses a "current residence" concept to determine who should be considered a resident of a sample household. Everyone who is currently living or staying at a sample address is considered a resident of that address, except people staying there for 2 months or less. People who have established residence at the sample unit and are away for only a short period of time are also considered to be current residents.

Custom Tabulations. The Census Bureau offers a wide variety of general purpose data products from the ACS. These products are designed to meet the needs of the majority of data users and contain predefined sets of data for standard census geographic areas, including both political and statistical geographic areas. These products are available through data.census.gov and the ACS Web site.

For users with data needs not met through the general purpose products, the Census Bureau offers Custom Tables on a cost-reimbursable basis, through the ACS Custom Tabulation program.[80] Custom tables are created by tabulating data from ACS microdata files. They vary in size, complexity, and cost depending on the needs of the sponsoring client.

Data.census.gov. The Census Bureau's primary tool for accessing population, housing, and economic data from the ACS, the Puerto Rico Community Survey, the decennial census, and many other Census Bureau data sets. Data.census.gov provides access to ACS data for a wide range of geographic areas, including states, cities, counties, census tracts, and block groups.

Data Profiles (DP). Data Profiles provide summaries of ACS data for various social, economic, housing, and demographic characteristics for the United States, regions, divisions, states, counties, county subdivisions, places, metropolitan areas, American Indian and Alaska Native areas, and other geographic areas. These profiles are similar in content to the demographic profiles from the decennial censuses. ACS Data Profiles begin with the letters "DP."

Decennial Census. The census of population and housing, taken by the Census Bureau in years ending in 0 (zero). Article I of the Constitution requires that a census be taken every 10 years for the purpose of reapportioning the U.S. House of Representatives. Title 13 of the U.S. Code provides the authorization for conducting the census in Puerto Rico and the Island Areas.

Derived Estimates. One of the benefits of working with ACS data is the ability to develop unique estimates called derived estimates. These derived estimates are usually based on aggregating estimates

[80] U.S. Census Bureau, American Community Survey (ACS), Custom Tables, <www.census.gov/programs-surveys/acs/data/custom-tables.html>.

across geographic areas or population subgroups for which combined estimates are not published in data.census.gov tables (e.g., aggregate estimates for a three-county area or for four age groups not collapsed).

Detailed Tables. Detailed Tables provide access to the most comprehensive ACS data tables on all topics and geographic areas. Tables include totals and subtotals. Users may choose more than one geographic area and more than one table that display in a scrolling list, but only what displays on the width of the screen will print.

ACS Detailed Tables begin with the letters "B" for base tables and "C" for collapsed tables. The "collapsed" tables cover the same topics as the base table, but with fewer categories.

Disclosure Avoidance. Statistical methods used before releasing data products to ensure the confidentiality of responses.

Estimates. Data for the ACS are collected from a sample of housing units and used to produce estimates of the actual figures that would have been obtained by interviewing the entire population using the same methodology.

Five-Year Estimates. Estimates based on 5 years of ACS data. These estimates are meant to reflect the characteristics of a geographic area over the entire 5-year period. These estimates are published for geographic areas down to the census block group level.

File Transfer Protocol (FTP). A process that allows a user to download large census and survey files and data sets from the Census Bureau's Web site.[81]

Geographic Comparison Tables (GCT). Allow users to compare ACS data across geographic areas in the same table (e.g., all counties in a state). ACS Geographic Comparison Tables begin with the letters "GCT."

Group Quarters (GQ) Facilities. A GQ facility is a place where people live or stay that is normally owned or managed by an entity or organization providing housing and/or services for the residents. These services may include custodial or medical care, as well as other types of assistance. Residency is commonly restricted to those receiving these services. People living in GQ facilities are usually not related to one another. The ACS collects data from people living in both housing units and GQ facilities.

Group Quarters (GQ) Population. Includes all people living in group quarters instead of housing units. Group quarters are places where people live or stay, in a group living arrangement that is owned or managed by an entity or organization providing housing and/or services for the residents. The group quarters population lives in group quarters, of which there are two general categories:

Institutional group quarters are facilities that house those who are primarily ineligible, unable, or unlikely to participate in the labor force while resident. The institutionalized population is the population residing in institutional group quarters, such as adult correctional facilities, juvenile facilities, skilled-nursing facilities, and other institutional facilities such as mental (psychiatric) hospitals and in-patient hospice facilities.

Noninstitutional group quarters are facilities that house those who are primarily eligible, able, or likely to participate in the labor force while resident. The noninstitutionalized population lives in noninstitutional group quarters such as college/university student housing, military quarters, and other noninstitutional group quarters such as emergency and transitional shelters for people experiencing homelessness and group homes.

Housing Unit. A housing unit is a house, an apartment, a mobile home or trailer, a group of rooms, or a single room occupied as separate living quarters, or if vacant, intended for occupancy as separate living quarters. Separate living quarters are those in which the occupants live separately from any other individuals in the building and which have direct access from outside the building or through a common hall. For vacant units, the criteria of separateness and direct access are applied to the intended occupants whenever possible.

Imputation. When data are missing, it is standard practice to use a statistical procedure called imputation to fill in missing responses. Imputation is the placement of one or more estimated answers into a field of data records that previously had no data or had incorrect or implausible data. There are two principal imputation methods to deal with missing or inconsistent data—assignment and allocation.

Margin of Error (MOE). The margin of error is the measure of sampling error published with each ACS estimate. A margin of error is the difference between an estimate and its upper or lower confidence bounds. Confidence bounds can be created by adding the margin of error to the estimate (for an upper bound) and subtracting the margin of error from the estimate (for a lower bound). All published margins of error for the ACS are based on a 90 percent confidence level.

[81] U.S. Census Bureau, American Community Survey (ACS), Data via FTP (File Transfer Protocol), <www.census.gov/programs-surveys/acs/data/data-via-ftp.html>.

Measurement Error. Also referred to as "response error," measurement error occurs when the response received differs from the "true" value as a result of the respondent, the interviewer, the questionnaire, the mode of collection, the respondent's record-keeping system(s), or other similar factors.

Multiyear Estimates. Three-year and five-year estimates based on multiple years of ACS data. ACS 5-year estimates are published for geographic areas down to the census block group level. ACS 3-year estimates have been discontinued, but are available for 2013 and earlier years for geographic areas with populations of 20,000 or more.

Narrative Profiles. Narrative profiles are short, analytic reports derived from recent ACS 5-year estimates. Each narrative profile covers 15 different topic areas and provides text and bar charts to display highlights of selected social, economic, housing, and demographic estimates for a selected geographic area.

Nonsampling Error. Total survey error can be classified into two categories—sampling error and nonsampling error. Errors that occur during data collection (for example, nonresponse error, response error, and interviewer error) or data capture fall under the category of nonsampling error.

Overcoverage. Overcoverage exists when housing units or people have more than one chance of selection in the sample, or are included in the sample when they should not have been.

Period Estimates. An estimate based on information collected over a period of time. For ACS estimates, the period is either 1 year or 5 years. ACS 3-year estimates have been discontinued, but are available for 2013 and earlier years for geographic areas with populations of 20,000 or more.

Point-in-Time Estimates. An estimate based on one point in time. The decennial census long-form estimates for the 2000 Census were based on information collected as of April 1, 2000.

Population Estimates Program. The Census Bureau's Population Estimates Program (PEP) produces July 1 estimates for years after the last published decennial census (2010), as well as for past decades. Existing data series—such as births, deaths, federal tax returns, Medicare enrollment, and immigration—are used to update the decennial census base counts. Population estimates are used in federal funding allocations, in setting the levels of national surveys, and in monitoring recent demographic changes.

Public Use Microdata Area (PUMA). A statistical area defined to contain a population of 100,000 or greater for which the Census Bureau tabulates Public Use Microdata Sample (PUMS) data. ACS and decennial census population and housing microdata are disseminated using these defined areas.

Public Use Microdata Sample (PUMS) Files. Computerized files containing a sample of individual records of people and households that responded to the ACS (stripped of all identifying information). The PUMS files permit analysis of specific population groups and custom variables that are not available through other ACS data products.

Puerto Rico Community Survey (PRCS). The counterpart to the ACS that is conducted in Puerto Rico.

Quality Measures. Statistics that provide information about the quality of the data from the ACS. Four different measures are provided with the annual data release: 1) initial sample size and final interviews, 2) coverage rates, 3) response rates, and 4) item allocation rates for all collected variables. Details are available in the technical documentation for the ACS products.

QuickFacts. A Census Bureau Web site that provides quick, easy access to facts about people, businesses, and geographic areas for all states, counties, and cities and towns with more than 5,000 people.

Ranking Tables. A table or product type that orders the states according to the numeric value of the data displayed. ACS Ranking Tables begin with the letter "R."

Reference Week. The calendar week preceding the date on which the respondents completed their questionnaires or were enumerated. This calendar week may not be the same for all people since the enumeration may not be completed in 1 week.

Reference Period. Time interval to which survey responses refer. For example, many ACS questions refer to the day of the interview; others refer to "the past 12 months" or "last week."

Residence Rules. The ACS uses a "current residence" rule to interview people who are currently living or staying in the sample housing unit as long as their stay at that address will exceed 2 months. See the entry on "Usual Residence" for information about residence rules in the decennial census.

Respondent. The person supplying survey or census information about his or her living quarters and its occupants.

Respondent Errors. The respondents' failure to provide the correct answer to a survey question for any reason, such as poor comprehension of the question meaning, low motivation to answer the question, inability to retrieve the necessary information, or an unwillingness to answer the question truthfully.

Sample. Entities selected for a specific survey.

Sample Data. Population and housing information collected on a continuous basis for selected areas in the ACS and other surveys where data are gathered from a selected group of respondents. No sample data were collected in the 2010 Census.

Sample Survey. A data collection activity involving observations or questionnaires for a sample of a population. These data are used to produce estimates for the entire population.

Sampling Error. Errors that occur because only part of the population is directly contacted. With any sample, differences are likely to exist between the characteristics of the sampled population and the larger group from which the sample was chosen.

Sampling Rate. Proportion of the addresses in a geographic area, or residents of a GQ facility, who are selected for interview in a particular time period.

Sampling Variability. Variation that occurs by chance because a sample of the population is surveyed rather than the entire population.

Selected Population Profiles. Selected Population Profiles can be used to show ACS data for a specific racial or ethnic group (for example, Alaska Natives), ancestry groups, or country of birth.

Single-Year (1-Year) Estimates. Estimates based on 1 year of ACS data. They are meant to reflect the characteristics of a geographic area over an entire 12-month period. ACS 1-year estimates are published for geographic areas with populations of 65,000 or more. Starting with the 2014 ACS, the Census Bureau is also publishing 1-year Supplemental Estimates—simplified versions of popular ACS tables for areas with at least 20,000 people.

Standard Error. The standard error is a measure of the deviation of a sample estimate from the average of all possible samples.

Statistical Significance, Test of. A test of statistical significance provides statistical evidence that indicates whether an observed difference between two estimates is likely due to chance ("not statistically significant"), or likely represents a true difference that exists in the population as a whole ("statistically significant"). Statistical significance in census data products is usually reported at the 90 percent confidence level.

Note that some statistical significance results displayed in data.census.gov may be based on unrounded estimates and standard errors and users may not be able to duplicate the results using the rounded estimates and measures of error as displayed in data.census.gov.

Subject Tables. Subject Tables include ACS data organized by subject area, providing an overview of the information that analysts most often receive requests for from data users. ACS Subject Tables begin with the letter "S."

Summary File. The ACS Summary File is a comma-delimited text file that contains all of the Detailed Tables for the ACS data releases.

Summary Level. Summary levels specify the content and hierarchical relationships of the geographic elements that are required to tabulate and summarize data. Each summary level has an assigned 3-digit summary level code to help programmers link each specific summary level to its appropriate use in a table, map, or other data summarization format. Some examples of summary levels are:

- 040: State
- 050: State-County
- 060: State-County-County Subdivision
- 150: State-County-Census Tract-Block Group
- 160: State-Place
- 314: Metropolitan Statistical Area-Metropolitan Division
- 430: Urban Area-State-County

It is important to distinguish between a summary level and a geographic area. A summary level represents the concept of a geographic level. For example, summary level 050, State-County, represents the concept of a county within a state. By comparison, a geographic area covers territory "on the ground," such as Madison County, Indiana.

Supplemental Estimates. Simplified Detailed Tables that provide access to the most recent ACS data at a lower population threshold than the standard 1-year tables. Available for selected geographic areas with 20,000 people or more. ACS Supplemental Estimates begin with the letter "K."

Undercoverage. The extent to which a sample does not include members of the target population thus preventing those members from having any chance of selection into the sample.

Unit Nonresponse. The failure to obtain the minimum required data from a unit in the sample.

Universe. The total number of units (e.g., individuals, households, or businesses) in the population of interest.

Usual Residence. Usual residence is a concept used in the decennial census to determine where a person should be counted in the census. Usual residence is defined as the place where a person lives and sleeps most of the time. This place is not necessarily the same as a person's voting residence or legal residence.

Variance Replicate Tables. These augmented ACS Detailed Tables include sets of 80 replicate estimates, which allow advanced users to calculate measures of error for estimates using the same methods that are used to produce the published margins of error (MOEs) in data.census.gov. These methods incorporate the covariance between estimates that approximate MOE formulas do not. They are published for a subset of the 5-year Detailed Tables and at selected summary levels. The 2010–2014 ACS 5-year estimates were the first for which these were available. We started by selecting "Block Group," then "Louisiana," "Orleans Parish, Louisiana," and "All Block Groups within Orleans Parish, Louisiana" (see Figure 6.2).