
American Community Survey Design and Methodology (January 2014)

Chapter 13: Preparation and Review of Data Products



[This page intentionally left blank]

Table of Contents

Chapter 13: Preparation and Review of Data Products	1
13.1 Overview	1
13.2 Geography	2
13.3 Defining the Data Products	3
13.4 Description of Aggregated Data Products.....	4
13.5 Public Use Microdata Sample (PUMS)	7
13.6 Generation of Data Products	7
13.7 Data Release Rules.....	8
13.8 Data Review and Acceptance.....	10
13.9 Custom Data Products.....	11

[This page intentionally left blank]

Chapter 13: Preparation and Review of Data Products

13.1 Overview

This chapter discusses the data products derived from the American Community Survey (ACS). ACS data products include the tables, reports, and files that contain estimates of demographic, social, economic and housing characteristics. These products cover geographic areas within the United States and Puerto Rico. The Public Use Microdata Sample (PUMS) files, which enable data users to create their own estimates, are also data products.

Most surveys of the population provide sufficient samples to support the release of data products only for the nation, the states, and possibly, a few substate areas. Because the ACS is a very large nationwide survey that collects data continuously every year, products can be released for many types of geographic areas, including many smaller geographic areas such as places, townships, and census tracts.

The first step in the preparation of a data product is defining the topics and characteristics it will cover. Once the initial characteristics are determined, they must be reviewed by the Census Bureau Disclosure Review Board (DRB) to ensure that individual responses will be kept confidential. Based on this review, the specifications of the products may be revised. The DRB also may require that the microdata files be altered in certain ways, and may impose restrictions on the publication of estimates for certain geographic areas based on the size of the ACS sample or the population size of such areas. These activities are collectively referred to as disclosure avoidance.

The actual processing of the data products cannot begin until all response records for a given year or years are edited and imputed in the data preparation and processing phases, the final weights are determined, and disclosure avoidance techniques are applied. Using the weights, the sample data are tabulated for a wide variety of characteristics according to the predetermined content. These tabulations are done for the geographic areas that have a sample size sufficient to support statistically reliable estimates, with the exception of 5-year period estimates, which are available for small geographic areas down to the census tract and block group levels. The PUMS data files are created by different processes because the data are a subset of the full sample data.

After the estimates are produced and verified for correctness, Census Bureau subject matter analysts review them. When the estimates have passed the final review, they are released to the public. PUMS estimates are reviewed in a separate process.

While the 2005 ACS sample was limited to the housing unit (HU) population for the United States and Puerto Rico, starting in sample year 2006, the ACS was expanded to include the group quarters (GQ) population. This step made the ACS sample representative of the entire resident population in the United States and Puerto Rico.

In 2007, 1-year period estimates for the total population and subgroups of the total population in both the United States and Puerto Rico were released for sample year 2006. Similarly, in 2008, 1-year period estimates were released for sample year 2007.

In 2008, the Census Bureau, for the first time, released products based on three years of ACS sample data collected from 2005 through 2007. In 2010, the Census Bureau released the first products based on five years of consecutive ACS samples, 2005 through 2009. Since several years of sample form the basis of these multiyear products, reliable estimates can be released for much smaller geographic areas than is possible for products based on single-year data.

In addition to data products regularly released to the public, the Census Bureau releases other data products developed through a fee-based special tabulations program. Government agencies, private organizations and businesses, or individuals can request special tabulations of data. As is the case for regular data products, special tabulation requests are reviewed before release by the Census Bureau's Disclosure Review Board (DRB) to assure protection of the confidentiality of individual responses. Section 13.6 provides information on methods used by the DRB to protect census data, including data from the ACS.

Chapter 14 describes the dissemination of the data products discussed in this chapter, the ACS data release schedule, and locations of various products and supporting documents on the Census Bureau's website.

13.2 Geography

The Census Bureau strives to provide ACS estimates for the geographic areas that are most important to data users. For example, data products reflecting ACS estimates are disseminated for many of the nation's legal and administrative areas, including states, American Indian and Alaska Native (AIAN) areas, counties, minor civil divisions (MCDs), incorporated places, congressional districts, as well as a variety of other geographic entities. Some geographic areas for which ACS estimates are produced, such as block groups, census tracts and census designated places, are defined and delineated in cooperation with state and local agencies. Others, such as urban areas, are defined and delineated by the Census Bureau, based on criteria developed by the Census Bureau and reviewed by federal, state, local, and tribal government data users.

Information on the types of geographic areas for which the Census Bureau publishes ACS estimates is available at:

https://www.census.gov/acs/www/data_documentation/areas_published/.

The publication of 1-, 3-, and 5-year ACS estimates is subject to limitations imposed by confidentiality restrictions. One- and 3-year estimates are also subject to data quality filtering, a process designed to limit the publication of low-reliability estimates. If a geographic area met the 1-year or 3-year threshold for a previous period, but dropped below it for the current period, it will continue to be published as long as the population does not drop more than five percent

Version 2.0 January 30, 2014

below the specific period threshold. The topics of confidentiality restrictions and data quality filtering are covered in more detail in Section 3.6.

The Puerto Rico Community Survey (PRCS) also provides estimates for legal, administrative, and statistical areas in Puerto Rico. The data release and threshold rules described above for the publication of 1-year, 3-year, and 5-year ACS period estimates for the U.S resident population apply for the publication of data for the PRCS.

Many areas for which ACS estimates are produced undergo periodic boundary changes due to annexations, detachments, or mergers with other areas. Each year, the Census Bureau's Geography Division, working with state and local governments, conducts the Boundary and Annexation Survey (BAS) to collect updated boundary information.¹ Minor corrections to the location of boundaries also can occur as a result of the Census Bureau's ongoing Master Address File (MAF)/Topologically Integrated Geographic Encoding and Referencing (TIGER®) Enhancement Project. The ACS estimates must reflect these legal area boundary changes, so all estimates are based on Geography Division resources that depict the geographic boundaries for legal areas as they existed on January 1 of the sample year. In the case of multiyear estimates, the boundaries of the entity for which the estimate is produced are those effective on January 1 of the final year of data collection for that multiyear estimate. For example, the boundaries of an entity for which a 2011-2013 estimate is produced are those in effect on January 1, 2013.

13.3 Defining the Data Products

For the 1999 through 2002 sample years, the ACS detailed tables were designed to be comparable with Census 2000 Summary File 3 to allow comparisons between data from Census 2000 and the ACS. However, when Census 2000 data users suggested certain changes to many tables, the Census Bureau implemented these preferences for the ACS products.

Once a preliminary version of the revised suite of products had been developed, the Census Bureau asked for feedback on the planned changes from data users (including federal agencies) via a *Federal Register* Notice (Fed. Reg. #3510-07-P). The notice requested comments on current and proposed new products, particularly on the basic concept of the product and its usefulness to data users. Data users provided a wide variety of comments, leading to modifications of planned products.

ACS managers determined the format of the new products for use in the 2005 ACS data release of data collected from the 2004 ACS. This made it possible for data users to become familiar

¹The BAS is described in more detail in the Glossary.

with the new products and to provide comments well in advance of the release of data for the 2005 ACS.

Similarly, a *Federal Register* Notice issued in August 2007 shared with the public plans for the data release schedule and products that would be available beginning in 2008. This notice was the first that described products for multiyear estimates. Again in 2009, a *Federal Register* Notice was issued to gather user feedback on the first ACS 5-year products.

Since 2010, an ACS Data Products Planning Working Group (DPPWG) and ACS Portfolio Management Governance Board (PMGB) have had a critical role in establishing more efficient processes for considering, reviewing, and approving proposals for new and modified data products. These groups have helped to allocate resources appropriately to develop new products, to consider innovative product design approaches suggested by both internal and external groups, and to revise existing products to make them more useful. They also help to ensure that technical constraints are fully assessed against technical requirements, and that potential product innovations align with future goals.

The ACS Data Users Group (ACS DUG) was formed in early 2013 to provide regional planners, demographers, community leaders and other data users with an externally managed forum for exchanging information on understanding and using ACS data. More information is available on ACS DUG activities at: <http://www.acsdatausers.org/>

13.4 Description of Aggregated Data Products

ACS data products can be divided into two broad categories: aggregated data products, and products representing extracts of the Public Use Microdata Sample (PUMS), described in Section 13.5.

Data for the ACS are collected from a sample of housing units (HUs), as well as the GQ population, and are used to produce estimates of the actual figures that would have been obtained by interviewing the entire population. The aggregated data products contain the estimates from the survey responses. Each estimate is created using the sample weights from respondent records that meet certain criteria. For example, the 2012 ACS estimate of people under the age of 18 in Chicago is calculated by adding the weights from all respondent records from interviews completed in 2012 in Chicago with residents under 18 years old.

This section provides a description of each aggregated product. Each product described is available as 1-year period estimates; unless otherwise indicated, they are also available as 3-year and 5-year estimates. The data products described below contain all estimates planned for release each year, including those from multiple years of data. Data release rules described in Section 3.6 will prevent certain 1-and 3-year period estimates from being released if they do not meet ACS requirements for statistical reliability.

Detailed Tables

The detailed tables provide basic distributions of characteristics. They are the foundation upon which other data products are built. These tables display estimates and the associated lower and upper bounds of the 90 percent confidence interval. They include demographic, social, economic, and housing characteristics, and provide 1-, 3-, or 5-year period estimates for the nation and states, as well as for counties, towns, and other small geographic entities such as census tracts and block groups, the latter of which are only available as five year estimates.

The Census Bureau's initial goal in defining ACS data products was to maintain a high degree of comparability between ACS detailed tables and Census 2000 sample-based data products. In addition, characteristics not measured in the Census 2000 tables were included in the ACS base tables. For example, the ACS 2012 data products include more than 1,470 detailed tables that cover a wide variety of characteristics and over 400 race and Hispanic origin iterations.

Data Profiles

Data profiles are high-level reports containing estimates for demographic, social, economic, and housing characteristics. For a given geographic area, the data profiles include distributions for such characteristics as sex, age, type of household, race and Hispanic origin, school enrollment, educational attainment, disability status, veteran status, language spoken at home, ancestry, income, poverty, physical housing characteristics, occupancy and owner/renter status, and housing value. The data profiles include a 90 percent margin of error for each estimate. Beginning with the 2007 ACS, the Census Bureau published a comparison profile that compares the sample year's estimates with those of each of the four previous years. For example, the 2012 comparison profile compared 2012 estimates with those of the 2011 ACS, 2010 ACS, 2009 ACS and 2008 ACS. These profile reports include the results of a statistical significance test for each previous year's estimate, compared to the current year. This test result indicates whether the previous year's estimate is significantly different (at a 90 percent confidence level) from that of the current year.

Narrative Profiles

Sourced from the data profiles, the narrative profiles allow users to explore a narrative and graphical presentation of the statistics for their own communities. This descriptive report describes a geographic area using custom text and graphics across fifteen topics. This product is available using the 5-year data and can be generated for the Nation, States (including Puerto Rico), Counties, Places, Metro/Micro Areas, Zip Code Tabulation Areas (ZCTAs), Tracts, and American Indian/Alaska Native (AIAN) Areas. The 5-year narrative profiles were released beginning in 2012 and are available at:

http://www.census.gov/acs/www/data_documentation/2012_narrative_profiles.

Subject Tables

Subject tables are similar to the Census 2000 quick tables, and, like them, are derived from detailed tables. Both quick tables and subject tables are predefined, covering frequently requested information on a single topic for a single geographic area. However, subject tables contain more detail than the Census 2000 quick tables or the ACS data profiles. In general, a subject table contains distributions for a few key universes, such as the race groups and people in various age groups, which are relevant to the topic of the table. The estimates for these universes are displayed as whole numbers. The distribution that follows is displayed in percentages. For example, subject table S1501 on educational attainment provides the estimates for two different age groups—18 to 24 years old and 25 years and older— as a whole number. For each age group, these estimates are followed by the percentages of people in different educational attainment categories (high school graduate, college undergraduate degree, etc.). Subject tables also contain other measures, such as medians, and they include the imputation rates for relevant characteristics. There are about 70 topic-specific subject tables released each year.

Ranking Products

Ranking products contain ranked results of many important measures across states. They are produced as 1-year products only, based on the current sample year. The ranked results among the states for each measure are displayed in tables and tabular displays that allow for testing statistical significance.

The rankings show approximately 90 selected measures. The data used in ranking products are pulled directly from a detailed table or a data profile for each state.

Geographic Comparison Tables

Geographic Comparison Tables (GCTs) contain the same measures that appear in the ranking products, plus additional 100 demographic measures that cannot be produced as ranking tables. They are produced as both 1-year and multiyear products. GCTs are produced for states as well as for substate entities, such as congressional districts. The results among the geographic entities for each measure are displayed as tables.

Selected Population Profiles

Selected Population Profiles (SPPs) provide certain characteristics from the data profiles for a specific race or ethnic group (e.g., Alaska Natives) or some other selected population group (e.g., people aged 60 years and older). SPPs are provided every year for 1- and 3-year estimates. In 2008, the requirement that SPPs could only be produced for sub-state areas of at least 1,000,000 persons was modified to allow for SPPs with minimum population sizes of 500,000. Another change to SPPs in 2008 was the expansion in the number of groups for which SPPs are produced to include additional country-of-birth groups. Groups too small to warrant an SPP for

a geographic area based on one year of sample data may appear in an SPP based on the 3-year accumulations of sample data.

Selected Population Tables and American Indians and Alaska Native Tables

The ACS Selected Population Tables and American Indians and Alaska Native (AIAN) Tables include a set of data products comparable to the Census 2000 Summary File 4/AIAN products. Representing a collection of detailed tables, data profiles, and subject tables, this product provides data iterated by detailed race, ethnicity, and tribal group. This product was first released in May 2012 and included estimates for the period 2006 to 2010. The Census Bureau plans to produce the Selected Population Tables and American Indians and Alaska Native Tables every five years.

Report Production Series and Analytical Applications

Subject matter analysts at the Census Bureau create analytical data products based on recent ACS data. The analytical products include reports, infographics, and online mapping applications. These products provide data users with insightful analysis on the current state of the nation. Some subject matter areas produce special tabulations of the data for their topics. The Census Bureau also produces methodological analysis products based on ACS collection and processing research. Information on these various analytic products can be found at the American Community Survey website at: http://www.census.gov/acs/www/library/by_year/2013/.

13.5 Public Use Microdata Sample (PUMS)

The PUMS comprises individual records that contain information collected about each person and housing unit (HU). PUMS files are extracts from the microdata for which disclosure avoidance technique are applied to protect confidential information about households or individuals. These extracts cover all of the same characteristics contained in the full microdata sample files. The only geography other than state shown on a PUMS file is the Public Use Microdata Area (PUMA). PUMAs are special non-overlapping areas that partition a state; each PUMA contains a population of about 100,000 or more. State governments drew the PUMA boundaries at the time of the Census. PDF-format maps of PUMA boundaries are available from the Census Bureau's website at: <http://www.census.gov/geo/maps-data/maps/reference.html>.

The Census Bureau releases 1-year, 3-year, and 5-year PUMS files. The multiyear PUMS files combine annual PUMS files to create larger samples in each PUMA, covering a longer period of time. This will allow users to create estimates that are more statistically reliable.

13.6 Generation of Data Products

The subject matter analysts in the Census Bureau's Social, Economic and Housing Statistics Division and Population Division provide the specifications for ACS data products. These

specifications include the logic used to calculate every estimate in each data product and the exact textual description associated with each estimate. Since 2006, limited changes to these specifications have occurred. The ACS Data Products Planning Workgroup, which serves as a Change Control Board, must approve specification changes. Once the specifications are validated and verified, tabulation starts when the weighted and edited microdata become available (see Chapters 10 and 11). ACS data products are generated by various automated tabulation systems, beginning with the 1-year period estimates for the detailed tables data products.

One distinguishing feature of the ACS data products system is that standard errors are calculated for all estimates and are released with the latter in tables. This practice differs from that followed for decennial census long form products released for Census 2000 and earlier decennial censuses, for which only the long form estimates appeared in data tables. Users of long form data for these decennial censuses were provided with a description and explanation of the concept of statistical uncertainty, and guidance for calculating the standard errors of individual estimates. However, the specific standard errors applying to each estimate were not calculated for data users and did not appear alongside the estimates in the data tables that included long form estimates.

Subject matter analysts use the standard errors in their internal reviews of estimates. Data users can also use them to determine whether two ACS estimates are statistically different.

13.7 Data Release Rules

Various kinds of restrictions are applied to ACS data to limit the disclosure of information about individual respondents and to limit the production of ACS estimates with unacceptable statistical reliability.

Population Thresholds for ACS Data

Population thresholds restrict the availability of data according to the type of ACS estimate. Areas or groups of 65,000 or more are eligible for 1-, 3-, and 5-year estimates. Areas or groups of 20,000 or more are eligible for 3- and 5-year estimates. Areas or groups of 20,000 or fewer are eligible for 5-year estimates.

Other population threshold rules apply in cases where a drop in population size changes the qualification of an area for types of ACS estimates for which it was previously qualified. Consider, for example, the case of a geographic area that met the 1-year threshold for a previous period, but dropped below it for the current period. The Census Bureau would like to provide the same kind of estimates for such areas every year to preserve the continuity of data for such areas. To give such an area a reasonable chance to retain access to 1-year estimates, the Census Bureau applies a data release rule that provides for the continued production of 1-year estimates unless the population drops more than five percent below the specific period threshold for 1 year estimates, 65,000. A comparable rule applies to areas that normally qualify for 3-year estimates.

Data Quality Filtering

Another kind of data release rule, data quality filtering, applies to ACS 1-year and 3-year estimates. Every detailed table consists of a series of estimates. Each estimate is subject to sampling variability that can be summarized by its standard error. If more than half of the estimates in the table are not statistically different from 0 (at a 90 percent confidence level), then the table fails to meet the rule's requirements and is restricted from publication. Dividing the standard error by the estimate yields the coefficient of variation (CV) for each estimate. (If the estimate is 0, a CV of 100 percent is assigned.) To implement this requirement for each table at a given geographic area, CVs are calculated for each table's estimates, and the median CV value is determined. If the median CV value for the table is less than or equal to 61 percent, the table passes for that geographic area and is published; if it is greater than 61 percent, the table fails and is not published.

Whenever a table fails, a simpler table that collapses some of the detailed lines together can be substituted for the original. If the simpler table passes, it is released. If it fails, none of the estimates for that table and geographic area are released. These release rules are applied to single- and 3- year estimates, but are not applied to the 5- year estimates.

Disclosure Review Board Rules

Beyond the population thresholds applied to ACS estimates and data quality filtering applied to ACS tables, the Disclosure Review Board (DRB) establishes additional rules that specify what ACS data are released. These rules describe, for example, how medians or other quantiles must be calculated, what requirements apply to means or aggregates, the thresholds for tables involving specific kinds of geographic areas or tables with more than 100 detail cells, and thresholds for tables of unweighted counts of people and housing units.

The DRB uses data swapping as a disclosure limitation technique for PUMS files. Data swapping involves the swap, or interchange, of a small percentage of household records between pairs of households in different geographic regions. The selection process for deciding which households should be swapped is highly targeted to affect the records with the most disclosure risk. Pairs of households that are swapped match on a minimal set of demographic variables, and the household pairs in most cases are located in the same census tract. All data products (tables and microdata) are created from the swapped data files.

For PUMS data the following techniques are employed in addition to swapping:

- Top-coding is a method of disclosure avoidance in which all cases in or above a certain percentage of the distribution are placed into a single category.
- Geographic population thresholds prohibit the disclosure of data for individuals or HUs for geographic units with population counts below a specified level.
- Age perturbation (modifying the age of household members) is required for large households containing 10 people or more due to concerns about confidentiality.

- Detail for categorical variables is collapsed if the number of occurrences in each category does not meet a specified national minimum threshold.

13.8 Data Review and Acceptance

After the editing, imputation, data products generation, disclosure avoidance, and application of the release rules are complete, subject matter analysts perform a final review of the ACS data and estimates before release. This final data review and acceptance process helps to ensure that there are no missing values, obvious errors, or other data anomalies.

Each year, the subject matter analysts review the ACS estimates following a multistep review process and provide clearance before estimates are released to the public. Because of the short time available to review such a large amount of data, an automated review tool (ART) and various multiyear review tools have been developed to facilitate the process. These tools enable subject matter analysts to detect statistically significant differences in estimates from one year to the next using several statistical tests. The initial version of ART was used to review 2003 and 2004 data. It featured predesigned reports and included functions allowing for ad hoc, user-defined queries for hundreds of single-year estimates covering 350 geographic areas. The improved version has been used by the analysts since June 2005. It was designed to work on much larger data sets and has a wider range of capabilities, with faster response time to user commands. In 2008, a team of programmers, analysts, and statisticians developed an automated multiyear review tool to assist analysts in their review of the multiyear estimates and this tool has been used for the review of 3-year and 5-year estimates.

The ACSO staff, together with the subject matter analysts, also have developed an automated tool to facilitate documentation and clearance for the edit review process. The edit management and messaging application (EMMA) is used to track the progress of analysts' review activities. EMMA reports are readily available to analysts, facilitating collaboration among them and contributing to a more efficient review process.

Important Notes on Multiyear Estimates

While the types of data products for the multiyear estimates are almost entirely identical to those used for the 1-year estimates, there are several distinctive features of the multiyear estimates that data users must bear in mind.

First, the geographic boundaries that are used for multiyear estimates are always the boundary as of January 1 of the final year of the period. Therefore, if a geographic area has gained or lost territory during the multiyear period, this practice can have a bearing on the user's interpretation of the estimates for that geographic area.

Secondly, for multiyear period estimates based on monetary characteristics (for example, median earnings), inflation factors are applied to the data to create estimates that reflect the dollar values in the final year of the multiyear period.

Finally, although the Census Bureau tries to minimize the changes to the ACS questionnaire, these changes will occur from time to time. Changes to a question can result in the inability to build certain estimates for a multiyear period containing the year in which the question was changed. In addition, if a new question is introduced in the middle of a multiyear period, estimates of characteristics related to the new question will not be available until they are included in the entire period.

13.9 Custom Data Products

The Census Bureau offers a wide variety of general-purpose data products from the ACS designed to meet the needs of the majority of data users. They contain predefined sets of data for standard census geographic areas. For users whose data needs are not met by the general-purpose products, the Census Bureau offers customized special tabulations on a cost-reimbursable basis through the ACS custom tabulation program. Custom tabulations are created by tabulating data from ACS edited and weighted data files. These projects vary in size, complexity, and cost, depending on the needs of the sponsoring client.

In some cases, requests for ACS custom tabulations from federal agencies with statutory requirements for information have led to innovative approaches in providing ACS data. For example, the Census Bureau met a Department of Justice tabulation request for ACS estimates to meet Section 203 requirements of the Voting Rights Act by providing modeled ACS estimates.

Each custom tabulation request is reviewed in advance by the DRB to ensure that confidentiality is protected. The requestor may be required to modify the original request to meet disclosure avoidance requirements. For more detailed information on the ACS Custom Tabulations program, go to: http://www.census.gov/acs/www/data_documentation/custom_tabulations/.