A Compass for Understanding and Using American Community Survey Data

What Researchers Need to Know

Acknowledgments

A Compass for Understanding and Using American Community Survey Data

What Researchers Need to Know

U.S. Department of Commerce
Gary Locke,
Secretary

Vacant,
Deputy Secretary

Economics and Statistics Administration
Vacant,
Under Secretary for Economic Affairs

U.S. CENSUS BUREAU
Thomas L. Mesenbourg,
Acting Director

ECONOMICS
AND STATISTICS
ADMINISTRATION

**Economics and Statistics
Administration**

Vacant,
Under Secretary for Economic Affairs

U.S. CENSUS BUREAU

Thomas L. Mesenbourg,
Acting Director

Thomas L. Mesenbourg,
Deputy Director and
Chief Operating Officer

Arnold A. Jackson
Associate Director for Decennial Census

Daniel H. Weinberg
Assistant Director for ACS and Decennial Census

Susan Schechter
Chief, American Community Survey Office

Contents

The American Community Survey (ACS) is a nationwide survey designed to provide communities with reliable and timely demographic, social, economic, and housing data every year. The U.S. Census Bureau will release data from the ACS in the form of both single-year and multiyear estimates. These estimates represent concepts that are fundamentally different from those associated with sample data from the decennial census long form. In recognition of the need to provide guidance on these new concepts and the challenges they bring to users of ACS data, the Census Bureau has developed a set of educational handbooks as part of *The ACS Compass Products*.

We recognize that users of ACS data have varied backgrounds, educations, and experiences. They need different kinds of explanations and guidance to understand ACS data products. To address this diversity, the Census Bureau worked closely with a group of experts to develop a series of handbooks, each of which is designed to instruct and provide guidance to a particular audience. The audiences that we chose are not expected to cover every type of data user, but they cover major stakeholder groups familiar to the Census Bureau.

General data users	Congress
High school teachers	Puerto Rico Community Survey data users (in Spanish)
Business community	Public Use Microdata Sample (PUMS) data users
Researchers	Users of data for rural areas
Federal agencies	State and local governments
Media	Users of data for American Indians and Alaska Natives

The handbooks differ intentionally from each other in language and style. Some information, including a set of technical appendixes, is common to all of them. However, there are notable differences from one handbook to the next in the style of the presentation, as well as in some of the topics that are included. We hope that these differences allow each handbook to speak more directly to its target audience. The Census Bureau developed additional *ACS Compass Products* materials to complement these handbooks. These materials, like the handbooks, are posted on the Census Bureau's ACS Web site: <www.census.gov/acs/www>.

These handbooks are not expected to cover all aspects of the ACS or to provide direction on every issue. They do represent a starting point for an educational process in which we hope you will participate. We encourage you to review these handbooks and to suggest ways that they can be improved. The Census Bureau is committed to updating these handbooks to address emerging user interests as well as concerns and questions that will arise.

A compass can be an important tool for finding one's way. We hope *The ACS Compass Products* give direction and guidance to you in using ACS data and that you, in turn, will serve as a scout or pathfinder in leading others to share what you have learned.

Introduction

Academic researchers have mined decennial census data on a variety of topics with geographic perspectives ranging from neighborhoods to the nation. The American Community Survey (ACS) is the new source for detailed demographic, social, economic, and housing characteristics. The Census Bureau is discontinuing the decennial long-form sample in the 2010 Census. For more information about the decennial census long form, review the text box on this page. The ACS is not simply a replacement, but rather a different kind of survey, utilizing continuous measurement approaches and a rolling sample. This handbook is aimed at three groups of academic researchers:

- Those who are familiar with the data files from the decennial census long form and are looking to continue their research with the ACS.

- Those who are familiar with using data—summary tabulations and microdata records—from complex sample surveys, although not necessarily decennial census data.

- Those who are learning about research methods and analysis of data from sample surveys—such as graduate students in methods classes. The materials in this handbook are no substitute for a methods course where analysis of secondary data from sample surveys is dealt with, along with the concepts of survey methodology and statistics for sampling. The handbook should, however, serve as an excellent supplementary text in such courses.

This handbook is intended to acquaint researchers with issues that may affect their analysis of ACS data. We have included case studies to illustrate uses of the ACS and exercises for readers to practice accessing and analyzing ACS data. The handbook does not cover all the issues of survey design and methodology associated with the ACS. The Census Bureau has produced a detailed summary of the methods used in the ACS

entitled, Design and Methodology: American Community Survey. It can be accessed at <http://www.census.gov/acs/www/Downloads /tp67.pdf>. This document provides a comprehensive discussion and is an essential reference for academic users of the ACS.

This handbook is supplemented with a set of technical appendixes. The appendixes vary in their level of methodological sophistication. Some readers may find one or two of the appendixes too elementary while others may find a few are too advanced. Throughout the handbook, we will refer to appendixes that provide expanded explanations of ACS issues. A glossary at the back of the handbook provides definitions of key terms.

The handbook focuses on the following topics relating to use of the ACS in academic research:

- Estimates for geographic areas
- Access to data products
- Sampling error
- Residence rules
- Income and poverty
- Migration
- Comparisons with Census 2000

Much of the academic research utilizing data from the decennial census long-form sample and now the ACS can be divided into three categories:

- An informational query for a table or graph to illustrate a point, a use that university researchers have in common with most users of census data. This includes looking for data tables and figures illustrating a point such as differences in the poverty rate or the percent in poverty that are recent migrants.

THE DECENNIAL CENSUS LONG-FORM SAMPLE

Every 10 years since 1790, Congress has authorized funds to conduct a national census of the U.S. population. The decennial census is required by the U.S. Constitution. Recent censuses have consisted of a "short form," which included basic questions about age, sex, race, Hispanic origin, household relationship, and owner/renter status, and a "long form." The long form was used at only a sample of households and included not only the basic questions on the short form but also detailed questions about social, economic, and housing characteristics. The questions on the long form supplied the raw data needed for a range of programs affecting education, veterans, employment, housing and community development, public health care, commuting, services for the elderly and disabled, and assistance programs for low-income families and children. About $300 billion in federal program funds are distributed each year based, in whole or in part, on these data.

What Researchers Need to Know **1**

- A query for records with specific attributes for geographic summary areas to create a data file with numerous observations that will be statistically analyzed. This involves data files for geographic areas, such as metropolitan areas, counties, places, minor civil divisions, tracts, or block groups, based on tabulations presented in summary files, such as Summary File 3 (SF3) from the Census 2000 long-form sample.

- A query for records with specific attributes for individual housing units, households, and people to create a data file with numerous observations that will be statistically analyzed. These are data files where the unit of analysis is the housing unit, household, or person extracted from the microdata files—either for the entire nation or for selected geographic areas.

Information Available From the ACS

The primary reason for the ACS is to help Congress determine funding and policies for a wide variety of federal programs. For this reason, the topics covered by the ACS are diverse. Nearly all topics in the ACS were also included on the Census 2000 long form. The ACS produces data on the demographic, social, and economic characteristics of the U.S. population and the physical and financial characteristics of the nation's housing. Table 1 lists the subjects included in the ACS. Many of the subjects contain numerous

characteristics. For example, "Journey to Work" includes data on means of transportation (automobile, bus, bicycle, walking), travel time (both duration and time departed), and whether or not a carpool is used. For more information about these subjects and the federal legislative and program uses of the data, see <http://www.census.gov/acs/www/Downloads /Final_2010_Census_and_American_Community _Survey_Subjects_Notebook.pdf>.

Table 1. Subjects Included in the American Community Survey

Demographic Characteristics
Age
Sex
Hispanic Origin
Race
Relationship to Householder
    (e.g., spouse)

Economic Characteristics
Income
Food Stamps Benefit
Labor Force Status
Industry, Occupation, and Class
    of Worker
Place of Work and Journey to
    Work
Work Status Last Year
Vehicles Available
Health Insurance Coverage*

Social Characteristics
Marital Status and Marital History*
Fertility
Grandparents as Caregivers
Ancestry
Place of Birth, Citizenship, and
    Year of Entry
Language Spoken at Home
Educational Attainment and
    School Enrollment
Residence One Year Ago
Veteran Status, Period of Military
    Service, and VA Service-
    Connected Disability Rating*
Disability

Housing Characteristics
Year Structure Built
Units in Structure
Year Moved Into Unit
Rooms
Bedrooms
Kitchen Facilities
Plumbing Facilities
House Heating Fuel
Telephone Service Available
Farm Residence

Financial Characteristics
Tenure (Owner/Renter)
Housing Value
Rent
Selected Monthly Owner Costs

*Marital History, VA Service-Connected Disability Rating, and Health Insurance Coverage are new for 2008.
Source: U.S. Census Bureau.

A Reengineered Census

The American Community Survey is one component of a reengineered decennial census and is designed to address the nation's need for more current information on the characteristics of its population and housing. Originally, the ACS was seen as a "replacement" for the long form. There was concern that the term "replacement" may have given users the impression that the parts were interchangeable, as in replacing a worn out set of spark plugs on your car. Rather, the change is more akin to switching from reliance on an internal combustion engine to a hybrid. On the outside, it looks the same and performs the same functions—only it is substantially different once you look under the hood and understand what provides the propulsion.

The ACS offers a number of benefits to researchers: timeliness, comparability, reliability, and numerous data products. The ACS data are timely and updated annually, compared with the decennial census long-form data that were released once a decade. The ACS uses uniform methods across the United States, enabling comparisons across geographic areas and over time. While the annual ACS sample size (about 2.5 percent) is smaller than the decennial long-form sample (about 16.7 percent), it is roughly 40 times larger than the Current Population Survey's Annual Social and Economic Supplement (March Supplement), which produces the nation's official estimates of poverty. The Census Bureau is preparing geographic area profiles, summary tabulations, and public use microdata files for each year's release of ACS data.

In order to deliver more timely information for all the geographic areas served by the decennial long form, the Census Bureau has designed the ACS as a sample survey using a continuous measurement approach to data collection. A sample of 3 million addresses is drawn from the Census Bureau's Master Address File (MAF) each year. For geographic areas with populations larger than 65,000, the sample is sufficient to produce reliable estimates based on a year's worth of responses. However, in order to provide estimates for areas with smaller populations, the sample must be accumulated over a number of years. The Census Bureau is producing 3-year estimates for areas down to populations of 20,000 or more and 5-year estimates for all units of census geography. This requires data users to understand the differences between 1-, 3-, and 5-year estimates—their availability and reliability. Refer to Appendix 1 for a more thorough explanation of multiyear estimates and Appendix 2 for greater detail on continuous measurement procedures and how they compare with "point-in-time" estimates.

Comparison of ACS and Decennial Census Sample Data

Differences between the design and methodology of the ACS and the decennial census can affect analyses of change based on data from the two surveys. A few of these differences are discussed here. For more information on this topic see:

- Appendix 2, "Differences Between ACS and Decennial Census Sample Data"

- Census Bureau's Web page "How to Use the Data: Guidance on Comparing ACS Data to Other Sources" <http://www.census.gov/acs/www/UseData/compACS.htm>

- Ten Things to Know About the American Community Survey (2005 Edition) <http://mcdc2.missouri.edu/pub/data/acs2005/Ten_things_to_know.shtml> by the Missouri State Data Center

Period and Point-in-Time Estimates

A major difference for users to understand is the change from estimates referenced to a specific date, as in April 1, 2000, for Census 2000, to ACS period estimates based on nearly daily collection of information over the entire calendar year. In practice, most of the decennial census socioeconomic sample data only approximate April 1 since the data are collected over a 6-month period and do not measure these characteristics as they existed on Census Day. For geographic areas of fewer than 65,000 people, it is necessary for the Census Bureau to accumulate 36 months of responses or even 60 months of responses prior to releasing ACS multiyear estimates. All ACS estimates are period estimates and are interpreted as the average values over the full time period. For example, 2007 ACS 1-year estimates describe the average characteristics for calendar year 2007 while 2005–2007 ACS 3-year estimates describe the average characteristics for the 3-year period of 2005, 2006, and 2007.

Residence Rules

Residence rules define who is included in a census or survey. The ACS does not use the same residence rule

as the decennial census. Rather than the decennial census rule of "usual place of residence," the ACS uses a 2-month residence rule. Residents who are living at the sample address at the time of the interview and who have lived or plan to live at that housing unit for more than 2 months are in scope for the survey. For areas with substantial seasonal populations, the difference in residence rules has the potential to affect characteristics such as occupancy rates, age composition, and household types. This is one of the clearest differences between the decennial census long-form data of the past, the upcoming short-form only 2010 Census, and the ACS. A detailed discussion of the impact of residence rules in general as well as differences between the ACS and the decennial census is presented in "Once, Only Once, and in the Right Place: Residence Rules in the Decennial Census (2006)," written by the Committee on National Statistics and published by the National Academies Press. Appendix 2 also addresses some of the key implications of these differences in residence rules for data users.

Vacant Units

The vacancy rate for an area where a high proportion of the housing units are owned by retirees could look different depending on the data source. Imagine a retired couple that occupies their second home for 4 months of the year—January through April. From the perspective of the local real estate market, the unit is not vacant because it is not for sale or rent. In the decennial census, it is not the "usual place of residence" for the couple and so even though they are living there on Census Day, April 1, the unit should be classified as "Vacant: For seasonal, recreational, or occasional use." If the same couple receives an ACS questionnaire during the time they are living there, and since they are "living or staying here for more than two months," the unit is classified as occupied. But if the unit falls into the sample during the other 8 months of the year, it is classified as vacant.

Age and Household Composition

The decennial census and the ACS may show different measures of the age of the population and the types of households in an area with a college or university. College students, whether living on or off campus, are considered to be residents of the area where they live while attending college—although they may not be year-round residents. The decennial census will include characteristics of the students in an area's age and household composition.

Because the ACS conducts ongoing interviews throughout the year, for 3 or more months of the year, it will interview many students where they are spending their summer, not in their college town. The ACS interviews are averaged over the collection period, thus the ACS

and the census may provide different pictures of the college town's characteristics.

Estimates of Income Characteristics

The Census Bureau cautions that income characteristics in the ACS and the decennial census may not be comparable because of some important differences in the universe and data collection. These differences are explained on the Census Bureau Web site at <http://www.census.gov/acs/www/UseData/Comparison_Guidance.htm#srcinc>.

While the Census Bureau designed the ACS questions, edits, and imputation procedures to be the same as those used in Census 2000, a major difference is the reference period. While Census 2000 asked for a respondent's income during 1999, the ACS collects data throughout the year on an on-going, monthly basis and asks for a respondent's income over the past 12 months. A paper presented by Census Bureau staff at the 2003 American Statistical Association Meetings reports on a comparison study of income data in Census 2000 and the 2000 ACS. The paper is available on the Census Bureau Web site at <http://www.census.gov/acs/www/Downloads/ACS/ASA_nelson.pdf>. Using a split panel test of the differences in reported income associated with the different reference periods, the investigators found income from wages and salary to be statistically higher for the panel asked to report on income for the previous calendar year when compared with income reported throughout the year. Income from self-employment was also significantly lower.

Data on income are collected from respondents as "Income in the past 12 months." Instructions to the respondent explain that the past 12 months refer to "the period from today's date one year ago up through today." See examples of ACS questionnaires at <http://www.census.gov/acs/www/SBasics/SQuest/SQuest1.htm> and Subject Definitions at <http://www.census.gov/acs/www/SBasics/SDef.htm>. This means that responses collected in January report income almost entirely from the calendar year prior to the reporting year and responses collected in December are for income received almost entirely in the reporting calendar year. The other months are a mix of income from the prior calendar year and the reporting calendar year.

In order to reduce the effect of inflation on the estimates of income, all income data are adjusted using the Consumer Price Index and stated in terms of annual dollars for the most recent year covered in the release. Only the income dollar amounts are adjusted for inflation in ACS single-year estimates. The household financial data (e.g., rent and mortgage payments, utility costs) are not adjusted when 1-year estimates are produced but are adjusted when 3- and 5- year esti-

mates are produced. Dollar values of income variables on the PUMS files, which consist of responses only for a single year are reported in nominal dollars and will need to be inflation-adjusted using the variable ADJUST. If the analyst combines PUMS files for different years, the dollar values for all financial variables should be adjusted to the latest year. Refer to Appendix 5 for more information on dollar-denominated data.

ACS estimates of income can be affected by the ACS methodology when a substantial portion of the population is not composed of year-round residents but rather includes a substantial proportion of those who live elsewhere for a portion of the year. The impact of the different residence rules used in the ACS on estimates of income for small areas will be greatest for areas with seasonal populations such as college towns, resort areas, and retirement communities.

Population Below the Poverty Level

The number and proportion of people living below the poverty threshold serve as indicators of the social and economic well being of an area. The Census Bureau produces a number of data sources that provide estimates of an area's population living below the poverty threshold. The Census Bureau's Housing and Household Economic Statistics Division maintains a Web site that provides users with details on the major national household surveys and programs that produce poverty statistics <http://www.census.gov/hhes/www /poverty/newguidance.html>. These major sources of estimates of poverty are:

- The American Community Survey (ACS)
- The Annual Social and Economic Supplement to the Current Population Survey (CPS ASEC)
- The Census 2000 long-form sample survey
- The Small Area Income and Poverty Estimates (SAIPE) program
- The Survey of Income and Program Participation (SIPP)

Differences in methodology for estimating the poverty rate among these sources can produce estimates that are not comparable. For example, there have been considerable changes in the SAIPE model-based estimates of poverty for counties when the Census Bureau switched from using poverty data measured by the CPS ASEC to data from the ACS as input to their model. A detailed explanation of the SAIPE estimation procedures is presented by the Census Bureau at <http://www.census.gov/hhes/www/saipe/methods /index.html>. As a second example, the Census Bureau provides a summary of key differences between the CPS ASEC and ACS poverty estimates at <http://www .census.gov/hhes/www/poverty/factsheet.html>.

Our focus for poverty status in this handbook is to illustrate comparisons between the ACS and Census 2000. The Census Bureau's ACS Web page on "How to Use the Data: Guidance on Comparing ACS Data to Other Sources" <http://www.census.gov/acs/www /UseData/compACS.htm> provides general guidelines regarding global differences between the 2006 and 2007 ACS and Census 2000 as well as guidance for specific items and subjects. The guidelines are available for all subjects in the ACS and are organized by subject area such as Age, Poverty, Housing, and Group Quarters. For example, two categories are included under the heading for Poverty: "Poverty Status of Families and People in Families" and "Poverty Status of All People in the Poverty Universe." The tables provide shorthand advice on comparisons that include "Compare," "Compare With Caution," and "Do Not Compare." In addition, there are links to "more info" and to crosswalks for comparable summary tables in the 2006 ACS and the Census 2000 Summary File 3 (SF3).

Geographic Mobility and Migration

In order to distinguish between movers and nonmovers, the ACS questionnaire asks, "Did this person live in this house or apartment one year ago?" Follow-up questions and response categories further classify individuals as local movers, domestic migrants, or immigrants from abroad. These data can be used to construct migration streams of people coming into and moving out of an area. For the out-migration stream, only those with a destination in the United States and Purto Rico can be identified. The Census Bureau produces summary tabulations and users have the option of constructing their own estimates—especially with detailed characteristics of subpopulations—using the ACS PUMS files. The ACS summary tabulations of mobility data that are available on the American FactFinder provide estimates of nonmovers, local movers, domestic in-migrants, and immigrants from abroad. Additional detailed tables providing estimates of the number of out-migrants are available beginning with the 2007 ACS data release.

The ACS data on mobility and migration are not comparable with the Census 2000 and earlier decennial data. The ACS asks about residence 1 year ago whereas Census 2000 asked about residence 5 years ago. The shorter reference period used in the ACS will do a better job of capturing short-term and subsequent return migration. Census 2000 captured longer-term migration. The proportion of the resident population that will be classified as migrants will generally be higher in Census 2000 than the ACS since it represents 5 years of accumulated migration instead of 1 year— minus return migrants that are undetected in a 5-year reference period. An excellent resource on the mobility and migration data from the 2006 ACS is the Census

Bureau's "2006 Subject Definitions" for the ACS and the Puerto Rico Community Survey (PRCS), available at <http://www.census.gov/acs/www/Downloads/2006/usedata/Subject_Definitions.pdf>. The section on "Residence One Year Ago" provides excellent information on the collection and processing of the data and the preparation of summary tabulations of the data. In addition, the publication covers limitations of the data and the history of the concept and question in prior versions of the ACS.

Kin Koerber, of the Census Bureau's Housing and Household Economic Statistics Division, produced a working paper examining domestic migration flows for states from the 2005 ACS (available on the Census Bureau Web site at <http://www.census.gov/acs/www/Downloads/State_Migration_Flows_Paper.doc>). Koerber states that while direct comparisons of migration data cannot be made between the ACS and Census 2000, it is possible to compare the relative size of flows. Koerber's paper could be useful to anyone who will be analyzing migration data from the ACS and in particular those who will be looking at trends involving decennial census data.

Sampling Error

Beginning with the 1940 Census, some characteristics reported from the decennial census have been estimates based on a sample of housing units and population. Sample estimates have variability associated with them, which is termed sampling error. Information about ACS sampling error is provided on the ACS Web site and measures of sampling error are included in the tables of tabulated characteristics. For Census 2000 and most other surveys, the documentation on sampling error is included in the detailed documentation for a file and users calculate margins of error on their own. Even academic researchers sometimes ignored the issues of sampling variability—which can be problematic when analyzing differences in small area estimates. Rather than taking into account that the estimates were derived from a complex sample survey, the estimates were sometimes treated as values for the population.

In the online presentation of ACS tables in the American FactFinder, the margins of error are presented alongside the estimates. The column heading "Margin of Error" is a hyperlink to the glossary that provides additional information about the margin of error, the confidence interval, and its lower and upper bounds. The margin of error presented in the American FactFinder is based on a 90-percent confidence level. Note that in the presentation of the margin of error prior to 2006, the Census Bureau used a factor of 1.65 in combination with the standard error of the estimate to derive the margin of error. In 2006 and subsequent years, the Census Bureau increased the precision by using a factor of 1.645.

By presenting the margin of error alongside the point estimates, users can more easily determine whether differences they observe over time and space are statistically significant or within the bounds of random variation. The Census Bureau uses a 90-percent

confidence level to determine the margin of error in the published tabulations. Depending on the application, a user may wish to increase the confidence level to 95 percent or 99 percent to conduct a more rigorous test of significant differences. Some users may combine geographic areas, thereby increasing the sample size and reducing the margin of error. Information about calculating the margin of error, addressing variations such as these, as well as examples, is included in Appendix 3.

Researchers using the microdata files need to calculate their own estimates of standard error due to sampling, using either the formula for generalized standard errors or by using the replicate weights. The Census Bureau uses the replicate weights method to estimate the standard error in the summary tabulations and also the table for "Tabulated 2006 PUMS Estimates for User Verification." The Census Bureau presents both approaches in Section 6, "Calculation of Standard Errors" of the document PUMS Accuracy of the Data (2006) <http://www.census.gov/acs/www/Downloads/2006/AccuracyPUMS.pdf>. According to the Census Bureau, "It is important to keep in mind that there will be differences between the standard error approximations computed by the two methods. Generally, using the replicate weights will produce a more accurate estimate of a standard error." Assuming that more users will opt for the formula to calculate generalized standard errors than will calculate standard errors using the replicate weights, the question becomes, "What difference does it make?" In Table 2, "Comparison of Two Approaches to Calculating Standard Errors in the ACS PUMS 2006," the two approaches in terms of Standard Error (SE), Margin of Error (MOE), and Coefficient of Variation (CV) for state-level estimates of the number of people by 5-year age categories are compared.

Basic Demographic Characteristics	Estimate	Replicate Weights			Generalized Standard Error		
		Standard Error (SE)	Coefficient of Variation (CV)	Margin of Error (MOE)	Standard Error (SE)	Coefficient of Variation (CV)	Margin of Error (MOE)
Total population*	19,306,183	NA	NA	NA	NA	NA	NA
Age 0–4	1,212,345	10,233	0.844%	16,833	12,727	1.050%	20,936
Age 5–9	1,204,832	11,762	0.976%	19,348	12,690	1.053%	20,875
Age 10–14	1,289,058	11,972	0.929%	19,695	13,096	1.016%	21,542
Age 15–19	1,391,848	9,190	0.660%	15,118	13,569	0.975%	22,321
Age 20–24	1,362,487	10,710	0.786%	17,617	13,436	0.986%	22,102
Age 25–34	2,509,761	13,017	0.519%	21,413	17,643	0.703%	29,023
Age 35–44	2,909,825	12,556	0.432%	20,655	18,770	0.645%	30,876
Age 45–54	2,827,340	11,639	0.412%	19,147	18,548	0.656%	30,512
Age 55–59	1,187,611	12,154	1.023%	19,994	12,605	1.061%	20,736
Age 60–64	892,536	10,784	1.208%	17,740	11,016	1.234%	18,122
Age 65–74	1,262,855	8,821	0.698%	14,510	12,971	1.027%	21,338
Age 75–84	892,184	8,355	0.936%	13,743	11,014	1.235%	18,118
Age 85 and over	363,501	6,970	1.917%	11,466	7,131	1.962%	11,730

*Total population estimates are controlled and therefore not subject to sampling error.
Adapted from "Tabulated 2006 PUMS Estimates for User Verification."

The formula for calculating generalized standard errors is below:

Total Formula: $SE(\hat{Y}) \doteq DF * \sqrt{99 * \hat{Y}\left(1 - \frac{\hat{Y}}{N}\right)}$

Where:
DF = Design Factor
N = Size of Geographic Area
$\hat{Y}$ = Estimate of Characteristic Total

Because the ACS sample is not a simple random sample, the design factor is used. The design factor approximates the increase in variance from a simple random sample that is due to the "actual sample design and estimation procedures used for the ACS" (page 9 of Accuracy of the Data). Reviewing the three sampling error measures displayed in the tables, there are very slight differences resulting from the method used to calculate the SEs. For example, the SE for the estimate of the population age 85 and over is 6,970 based on replicate weights and 7,131 based on generalized standard errors. The CV for this estimate is therefore higher for the generalized standard error approach (1.962 percent versus 1.917 percent). There is a tendency for the formula for generalized standard errors to estimate a higher standard error than the replicate weights method. This means that by using the generalized formula instead of the replicate weights, one is less likely to find statistically significant differences where they actually do exist. The lesson to take away from this is the necessity to learn how to use the more complex method with replicate weights to produce estimates of standard error.

Data Release and Geographic Areas

Tabulations are prepared based on accumulated responses to the survey questionnaire. Depending on the population size of a geographic area, tabulations will be based on the responses accumulated for 1-year, 3-year, or 5-year periods. Decennial long-form data were based on a "once a decade" 1-in-6 sample of housing units (with variations for rural and high-density urban areas), and all data products flowed from that collection period. ACS data are currently based on an annual sample that is approximately 1-in-40 housing units (with some variations), administered as 12 independent monthly samples, and data products are produced for geographic areas once enough responses have been accumulated (see Table 3). Data for geographic areas with populations of 65,000 and more are summarized and produced annually based on the accumulated responses from the previous calendar year's data collection. Areas with populations of 20,000 and more are summarized and produced annually based on the accumulated responses to the previous 3 years of data collection. All size areas are summarized with data produced annually based on the accumulated responses to the previous 5 years of data collection.

Table 3. **Release Schedule for ACS Data**									
		Year of Data Release							
Data product	**Population threshold**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**
		Year(s) of Data Collection							
1-year estimates	65,000+	2005	2006	2007	2008	2009	2010	2011	2012
3-year estimates	20,000+			2005–2007	2006–2008	2007–2009	2008–2010	2009–2011	2010–2012
5-year estimates	All areas*					2005–2009	2006–2010	2007–2011	2008–2012

*Five-year estimates will be available for areas as small as census tracts and block groups.
Source: U.S. Census Bureau.

The Census Bureau is producing three sets of tabulations for geographic areas with a population size of 65,000 and more: 1-year, 3-year, and 5-year estimates. The tabulations based on 5 years of data collection will have the smallest level of sampling error due to the larger sample size, that is, 5 years of accumulated responses. Two sets of tabulations are prepared for areas with populations of 20,000 and more. There will be only one set of small area estimates for areas with populations less than 20,000—which includes smaller counties and places along with census tracts and block groups. For these multiyear estimates, the records are pooled in a single data set and reweighted to a new set of controls based on the most current population estimates for each of the single years. For summary areas where boundaries change during the collection period—such as municipalities that may annex surrounding territory—the multiyear estimates are retabulated to fit the boundaries of the area as it existed for the most recent year's data. The individual housing and person records are also reweighted.

When comparing smaller areas with larger areas, researchers need to keep in mind which tabulations they are using and should use consistent collection periods for comparison purposes. For example, when analyzing census tract characteristics for a major metropolitan area, data users should use the 5-year estimates for both the tract-level data (your only option) and the metropolitan area.

Users of the ACS Public Use Microdata Samples (PUMS) will have access to the same geographic areas that were available from the Census 2000 PUMS. These Public Use Microdata Areas, or PUMAs, are the smallest geographic unit identified in the public use files. PUMAs for both the ACS and Census 2000 contain a minimum of 100,000 people. Maps for PUMAs can be found on the Census Bureau's Web site at <http://www.census.gov/geo/www/maps/puma5pct.htm>.

Access to Data Products

A summary of the current ACS data products is available on the Census Bureau's Web site at <http://www.census.gov/acs/www/Products/>. Most ACS data products are accessed from the American FactFinder—a data dissemination tool on the Census Bureau's Web site. This section discusses two types of products of major interest to academic researchers, Summary Files and PUMS files. The Summary Files contain tabulations of data collected in the ACS presented as a series of tables with frequency distributions for a single characteristic, such as age, or cross-tabulations of two or more characteristics, such as age by sex by race. The records in the Summary Files refer to a specified geographic area, such as a state, county, place, or census tract. Public use microdata files are like computerized versions of the completed and edited questionnaires—minus any personal identifying information. There are separate records for the housing units followed by individual records for each of the persons occupying the housing units.

ACS Summary Files

The ACS Summary Files are analogous to Summary File 3 or SF3 from Census 2000. The data files and documentation are available from the Census Bureau's ACS

Figure 1. **Screenshot of ACS FTP Site**

Source: U.S. Census Bureau, accessed at <http://www.census.gov/acs>.
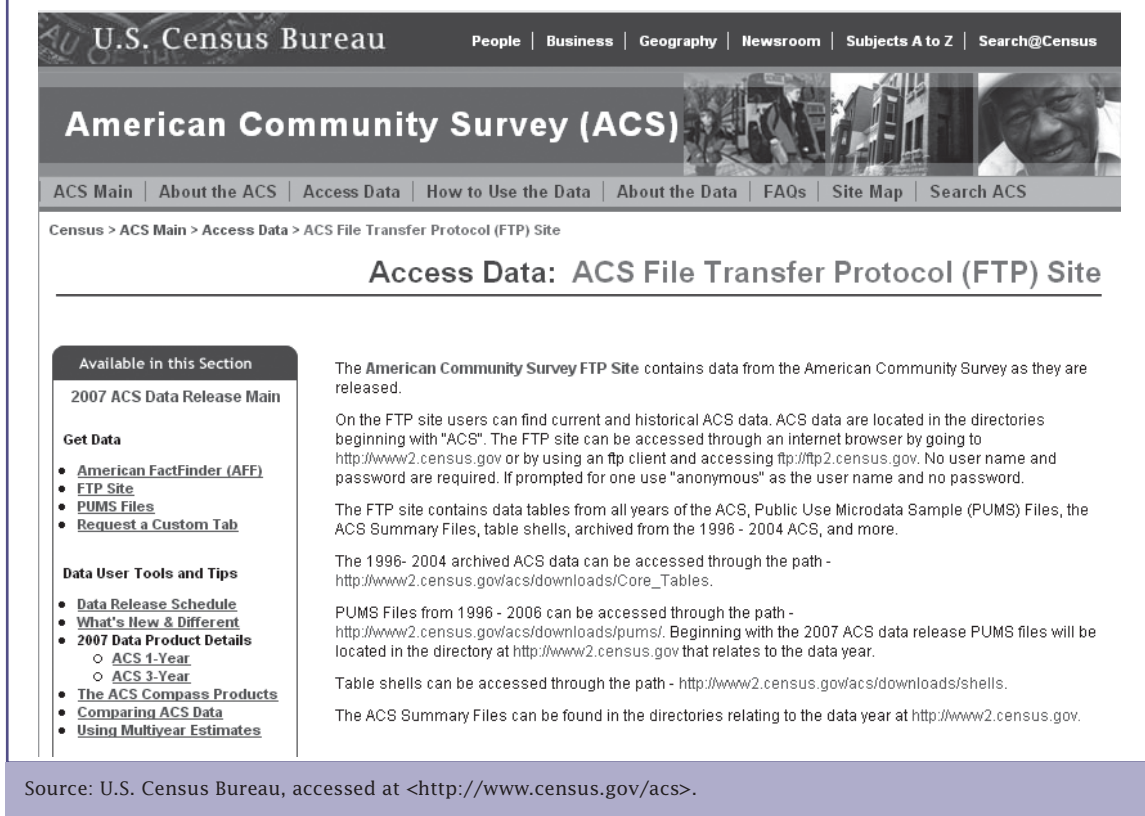
FTP site at <http://www.census.gov/acs/www /Special/acsftp.html>. See Figure 1 for a display of the ACS FTP site. The 2006 ACS Summary File Technical Documentation provides users with all the information they need to access and process these data, including survey methods and links to sample SAS programs for processing the data files. Users should note that new ACS Summary Files are released each year, including recently released 2007 ACS 1-year Summary Files and 2005–2007 ACS 3-year Summary Files. The documentation enables users to take advantage of the links to files on the Census Bureau Web site. The structure of the ACS Summary Files is slightly more complex than Census 2000 Summary Files because of the additional information on the reliability of sample estimates. Files that are merged with the ACS estimates provide information on the margin of error and the standard error.

ACS Public Use Microdata Sample Files

The ACS PUMS files are very similar to the PUMS files from Census 2000. Links to the data files and documentation are available from the Census Bureau's ACS Web site by selecting the link to PUMS <http://www .census.gov/acs/www/Products/PUMS/>. This Web page summarizes basic information about the ACS PUMS and provides necessary links to the data and documentation as illustrated in Figure 2. ACS data users

interested in more information about this resource can also refer to the PUMS Handbook that is part of the *ACS Compass Products* <http://www.census.gov/acs/www /UseData/Compass/compass_series.html>.

There are a number of reasons why academic researchers use the microdata files. Microdata provide the flexibility to create custom tabulations or to investigate the relationship among characteristics captured by the survey questionnaire. For research aimed at understanding small population groups, analyzing individuals, households, or housing units, and not requiring small area characteristics, the PUMS files serve the need.

The ACS PUMS files are organized as hierarchical files, in the same manner as the Census 2000 PUMS files. There are separate records for housing units and population. The housing unit records have unique identifiers that are repeated on each of the population records for people living in that housing unit. In this manner, housing unit characteristics can be appended to population records as needed in an analysis. For example, housing unit records contain variables providing the geographic location of the unit, and in order to identify a person's geographic area of residence, it is necessary to link that information from the housing unit record to the population record.

Source: U.S. Census Bureau, accessed at <http://www.census.gov/acs/www/Products/PUMS/>.

Because the records in the file represent a sample of the population, each record—housing and person—is assigned a weight. The weight is a numeric variable expressing the number of housing units or people that an individual microdata record represents. The sum of the housing unit and person weights for a geographic area is equal to the estimate of the total number of housing units and people in that area. Since the ACS is not a simple random sample survey but rather, a complex sample survey, the values of the weights vary. In order to generate estimates of the population based on the sample records, it is necessary to correctly use the weights assigned to each of the records. The Census Bureau provides basic tabulations of weighted characteristics from the ACS PUMS that researchers can use to verify the accuracy of their programming.

On the ACS PUMS Web page at <http://www.census.gov/acs/www/Products/PUMS>, there are files in various formats under the heading "Tabulated 2006 PUMS Estimates for User Verification" for the user to verify they are accessing and processing the PUMS files correctly. Unlike the Current Population Survey (CPS) or the American Housing Survey (AHS), where the microdata files are the full files and produce the same results as the printed reports, the ACS PUMS files are a subsample drawn from the full sample. The ACS PUMS file is a systematic subsample of the full ACS responses, designed to yield an effective sampling rate of 1 percent of each state's estimated population. Just like the Census 2000 PUMS, tabulations from the ACS PUMS will not match the summary tables from the American FactFinder. In order to verify that you have correctly accessed and tabulated data from the ACS PUMS file, it is recommended that you first replicate the values presented in "PUMS Estimates for User Verification."

Case Studies

Case Study 1: Interpreting Small Area 5-Year Period Estimates

Purpose: *Multilevel Analysis of Neighborhood Well Being*

Data File: *2001–2005 ACS 5-Year Estimates From the Multiyear Estimates Study*

Geographic Areas: *Counties and Tracts*

Background

This case study demonstrates how to use ACS data when you are working with different size geographic areas. It highlights the comparisons that should (and should not) be made. If you want to know how unemployment and poverty rates for census tracts compare with those rates for their county or metropolitan area, you need to work with ACS multiyear estimates. The ACS estimates for census tracts are only produced as ACS 5-year (60-month) period estimates and will be released for the first time for all census tracts in 2010 based on responses collected during the 2005–2009 calendar years.

The ACS provides estimates of characteristics, such as poverty and unemployment that can be used as indicators of well-being. Census tracts are frequently used to represent neighborhoods while a county or metropolitan area may serve as a geographic unit to represent a regional labor market. The ACS estimates can be used for a multilevel analysis of the importance of regional labor market conditions upon neighborhood levels of well-being. If we believe that neighborhood conditions are in part a function of regional conditions, then in examining the differences in levels of well-being (such as rates of poverty and unemployment), we would want to control our analysis of tracts by characteristics of the county or metropolitan area.

Assuming that all the labor market regions we are analyzing—represented by individual counties, groups of counties, or metropolitan areas—are areas with populations greater than 65,000, then summary tabulations will be available in the form of 1-year, 3-year, and 5-year estimates. For census tracts, only the 5-year estimates will be available. This case study will illustrate which estimates of county characteristics should be used in this analysis.

In December 2008, the Census Bureau released ACS 3-year estimates on the American FactFinder. The Census Bureau will release 5-year estimates in 2010. Therefore, the 5-year estimates used in this case study are not available from the American FactFinder.

However, the Census Bureau chose 36 counties as test sites for the development of the ACS and began collecting data in these counties prior to 2000. The Census Bureau released 1-, 3-, and 5-year estimates for these areas as part of the Multiyear Estimates Study. Researchers may access these data, as well as other useful information about the Census Bureau's Multiyear Estimates Study at <http://www.census.gov/acs /www/AdvMeth/Multi_Year_Estimates/overview.html>. Using this link, 1-, 3-, and 5-year estimates can be accessed for 34 of the test counties.

Analysis

In this case study two counties are considered—Pima County, Arizona, and Rockland County, New York. Specifically we will analyze the 1-year and 5-year estimates of unemployment for selected tracts and compare them with the unemployment rates at the county level. Unemployment rates are displayed in Table 4 for Pima County, Arizona, Rockland County, New York, and selected census tracts in these counties to illustrate the options that are available. Both counties have populations of 65,000 or more and therefore they qualify for 1-year, 3-year, and 5-year estimates. The county-level data for 5-year and 1-year period estimates are presented. The only tract-level data available are 5-year estimates for the 2001–2005 period. The most current estimates in Table 4 are the 1-year estimates for 2005. Those estimates, however, are not available for census tracts. Although the impulse to use the most current estimates is strong, the proper county-level estimates to use in the multilevel analysis are the 5-year estimates for the period 2001–2005. This means that an unemployment rate of 2.3 percent for tract 112.00 in Rockland County, New York, should be compared with the 5-year estimate of the Rockland County rate of 3.1 percent.

The Multiyear Estimates Study includes 5-year estimates for 1999–2003, 2000–2004, and 2001–2005. Trend analysis with the multiyear estimates should be done using estimates from nonoverlapping periods. For example, if you compare the 2000–2004 estimates with the 2001–2005 estimates you would be using the same observations for four of the component years. That is, the responses collected in 2001, 2002, 2003, and 2004 are common to both sets of estimates. The optimal comparison would be between the 2000–2004 estimates and the 2005–2009 estimates. Analysis of overlapping periods is possible but requires special attention to several factors. Appendix 4 discusses comparisons—including those involving overlapping periods—in greater detail.

Table 4. **American Community Survey Tabulations of Unemployment Rate**												
	5-Year Estimate		**1-Year Estimates**									
	2001-2005		**2001**		**2002**		**2003**		**2004**		**2005**	
	Estimate	**MOE**	**Estimate**	**MOE**	**Estimate**	**MOE**	**Estimate**	**MOE**	**Estimate**	**MOE**	**Estimate**	**MOE**
Pima County	4.1%	+/-0.2	3.6%	+/-0.3	4.5%	+/-0.4	4.5%	+/-0.4	3.6%	+/-0.4	4.2%	+/-0.4
Selected Tracts												
40.59	1.7%	+/-1.0		NA		NA		NA		NA		NA
44.16	2.2%	+/-1.1										
40.60	2.6%	+/-1.2										
47.12	3.9%	+/-2.6										
40.51	4.6%	+/-2.6										
40.61	6.0%	+/-5.1										
19.00	7.1%	+/-4.8										
7.00	8.1%	+/-3.6										
43.22	9.5%	+/-4.1										
39.02	13.0%	+/-6.4										
51.00	26.4%	+/-11.5										
	3.1%	+/-0.3	2.5%	+/-0.4	2.7%	+/-0.5	4.3%	+/-0.7	3.5%	+/-0.6	2.8%	+/-0.6
Rockland County												
Selected Tracts												
112.00	2.3%	+/-1.7		NA		NA		NA		NA		NA
105.02	2.9%	+/-1.5										
101.02	3.2%	+/-2.3										
105.01	3.7%	+/-1.7										
134.00	4.3%	+/-2.5										
115.01	5.0%	+/-2.4										
130.02	5.6%	+/-2.8										
115.04	6.4%	+/-3.1										
122.02	7.2%	+/-3.7										
113.01	8.2%	+/-3.9										
118.00	11.9%	+/-5.5										

Source: U.S. Census Bureau, American Community Survey Multiyear Estimates Study,
<http://www.census.gov/acs/www/AdvMeth/Multi_Year_Estimates/online_data_year.html>.

Conclusion

This case study reminds researchers that they should make comparisons using the same data series. The 1-year estimates should only be compared with other 1-year estimates. It may be necessary to use less current data for a given area in order to make appropriate comparisons. Researchers should also avoid comparing ACS estimates for overlapping time periods.

Case Study 2: Estimates of Income

Purpose: Compare ACS and Quarterly Census of Employment and Wages (QCEW) Data on Earnings
Data Files: ACS 1-Year Estimates for 2000–2006, QCEW Annual Earnings, 2000–2006
Geographic Areas: County

Background

This case study illustrates how the ACS estimates of income from wages and salaries compare with other sources of income data. We use the Tucson, Arizona, metropolitan area for this example. It comprises one county, Pima County, which is a major retirement and seasonal destination as well as home to the University of Arizona. Pima County, Arizona, was a test site for the development of the ACS and therefore a series of 1-year estimates are available from the Multiyear Estimates Study for 2000 through 2006. Data on wages from the Bureau of Labor Statistics' Quarterly Census of Employment and Wages (QCEW) are to be compared with the ACS estimates of aggregate wage or salary income for the metropolitan area. The ACS is a household survey and the data on earnings are based on place of residence, while the QCEW is based on administrative records filed by employers that are based on place of work. The Census 2000 County-to-County

Worker Flow Files reported that 97.5 percent of those working in the Tucson, Arizona, metropolitan area lived there as well, which minimizes the differential in earnings based on place of residence versus place of employment. The QCEW data on wages are for workers covered by unemployment insurance, while the ACS is for all workers.

Analysis

Table 5 and Figure 3 display the ACS 1-year estimates of wage and salary income for the Tucson, Arizona, Metropolitan Statistical Area (MSA) for the years 2000 to 2006. As noted earlier, estimates of income characteristics in the ACS are split across the reporting year and prior year. To make the most reasonable comparison, we have averaged the QCEW data over 2 years and "synched" them to the ACS reporting year. The ACS data show higher salary and wage levels than the QCEW data—even taking the margins of error into account. This is consistent with the more complete coverage in the ACS of workers who are not covered by unemployment insurance. Wage and salary income for 1999 as reported in Census 2000 was also higher than income from the QCEW for 1999. The trend line for ACS data is not as smooth as the QCEW trend line, which is consistent with most comparisons of data from a sample survey and an enumeration.

Conclusion

This case study highlights the need for researchers to understand that different sources of data will provide different measures for specific topics, such as income. It is important to spend the time understanding the data sources before you use them, to understand which source is the best source for your needs. In this example, the major difference in the two data sets is that the ACS is a sample survey of households while the QCEW is an enumeration of employers to collect data on employees covered in the unemployment insurance program.

Table 5. **Comparison of Wage and Salary Income for Tucson, AZ MSA From the ACS and QCEW**			
	ACS		
	Estimates (in dollars)	MOE (in dollars)	QCEW* (in dollars)
2000	10,650,578	222,164	9,269,883
2001	10,961,591	235,276	9,810,476
2002	12,140,897	369,319	10,177,163
2003	12,430,720	372,496	10,540,859
2004	12,721,230	374,256	11,215,675
2005	14,244,673	408,409	12,133,677
2006	15,188,623	424,550	13,121,882

Source: U.S. Census Bureau, American Community Survey.   U.S. Bureau of Labor Statistics, Quarterly Census of Employment and Wages.
*QCEW data have been averaged across 2 years—the reporting and prior year.

Figure 3. **Comparison of Earnings Data for Tucson MSA From ACS and QCEW**
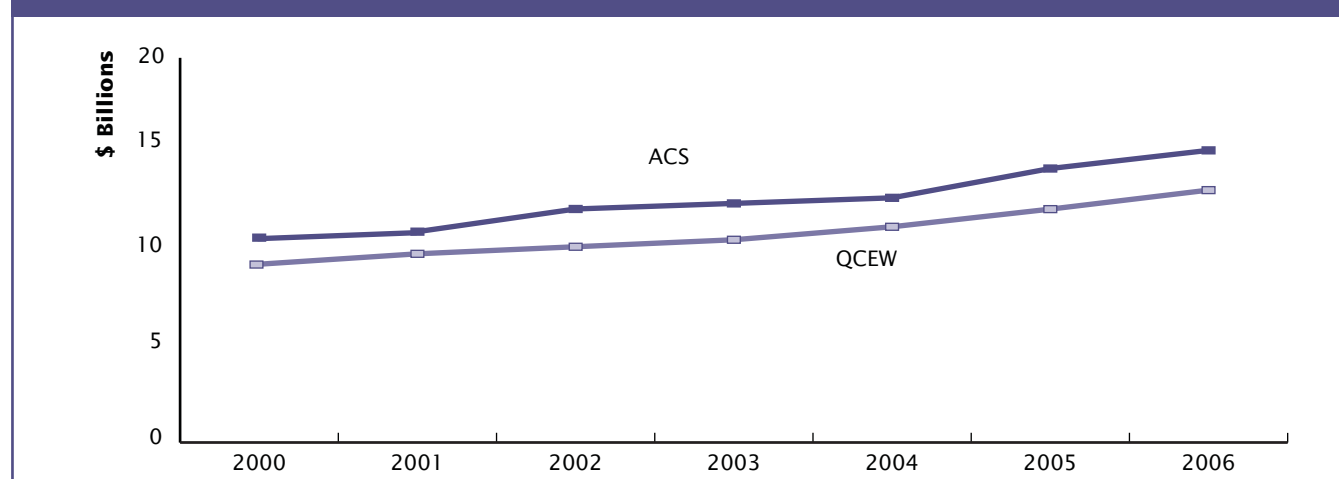
Source: U.S. Census Bureau and Bureau of Labor Statistics.

Case Study 3: Trends in the Poverty Rate

Purpose: Compare Poverty Rates Over Multiple Years

Data Files: ACS 1-Year Estimates, Small Area Income, and Poverty Estimates (SAIPE) Annual Estimates

Geographic Areas: County

Background

The following case is based on a newspaper article examining trends in the poverty rate for a county with a large college student population. The data points have been expanded to illustrate differences in the poverty rate from difference sources—especially when comparing trends over time. With the release of profiles from the 2005 ACS, journalists were mining the reports for newsworthy items. Tompkins County, New York, is home to Cornell University and approximately 10 percent of the county's population is college students in dormitories. The Ithaca Journal reported, "In Tompkins County, 20.7 percent were living in poverty last year, a jump from 12.9 percent in 2004," (Ithaca Journal, 8/30/2006) citing the source simply as U.S. Census Bureau. There were two mistakes in the article. First, the estimate of a 12.9 percent poverty rate was not for 2004 but for 2003. Second, data from the Census Bureau's Small Area Income and Poverty Estimates (SAIPE) for 2003 should not have been compared with the newly released 2005 ACS data. Figure 4 provides a more complete picture of the trend in poverty rates for Tompkins County, New York, from these data

sources than was available in August 2006. It demonstrates the extent to which estimates of the poverty rate are not comparable.

Analysis

Figure 4 shows the SAIPE series of estimates along with the estimates from Census 2000 and the ACS. SAIPE is the product of model-based estimates that used the Census Bureau's Current Population Survey (CPS) as a major input through 2004. For 2005, SAIPE was based on the ACS, creating an abrupt break in the series between 2004 and 2005 for Tompkins County. The Census Bureau informed SAIPE users of this change in methodology and highlighted cautions relating to comparisons of 2004 to 2005 results. See <http://www.census.gov/hhes/www/saipe/methods/cautions.html>.

Conclusion

All of the estimates in Figure 4 are "right," but they are based on different data sources. There are important factors to consider if a newspaper wanting to compare annual fluctuations in the poverty rate for a county contacted you. You would need to decide which data to recommend. You should avail yourself of the reference materials provided on the Census Bureau's Web site. This case study also demonstrated the need to be clear about when it is and is not appropriate to compare data from two different sources. In this example, comparisons of SAIPE with ACS data are inappropriate.

Figure 4. **Trends in Poverty Rate and Noncomparable Sources—Tompkins County, NY**
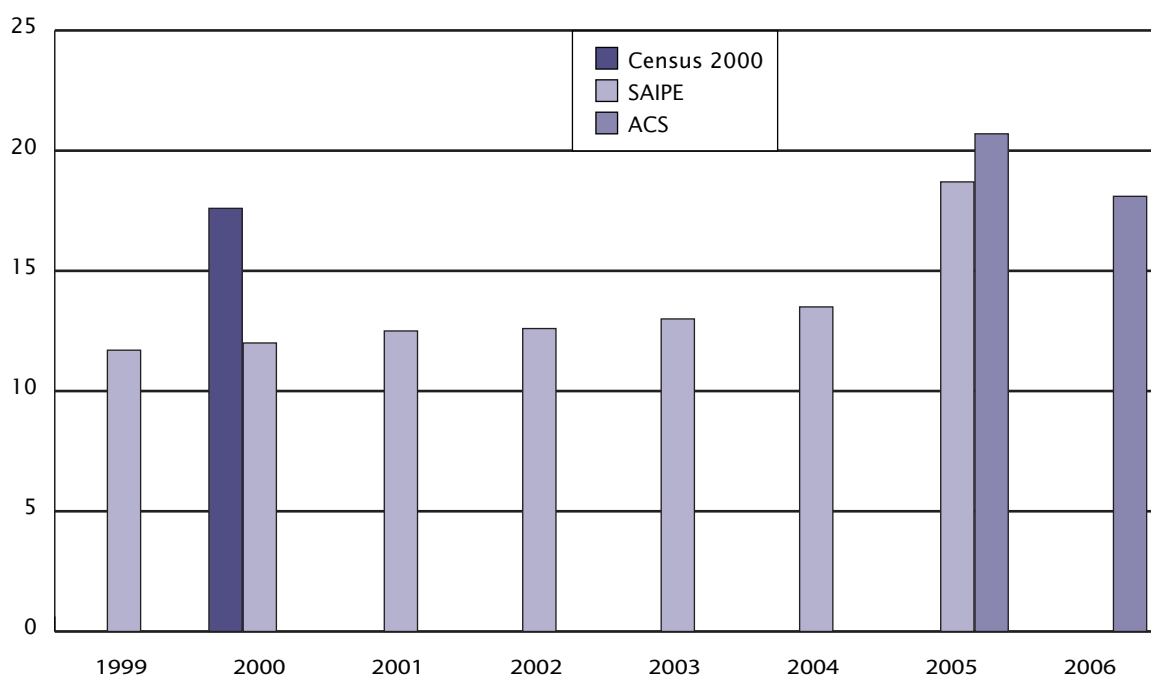
Source: U.S. Census Bureau.

Exercises

Exercise 1: Access Summary Tabulations

Task: Access Detailed Poverty Data From ACS Detailed/ Summary Tables

Data File: 2006 American Community Survey

Geographic Area: States

Background

Accessing detailed tabulations from the ACS for a number of geographic areas can be accomplished in different ways. This exercise provides an example using the American FactFinder to access 2006 ACS data on poverty status for all states. One could also use the American FactFinder to access poverty data for Core Based Statistical Areas (which include Metropolitan and Micropolitan Statistical Areas), counties, or a wide variety of census geographies meeting the population threshold of 65,000 for 1-year estimates. This exercise assumes a basic familiarity with using the American FactFinder interface to access census data.

Accessing and Downloading the Data

This section details the steps required to access the data of interest. It provides screen shots of some of the Web pages you will encounter. First go to the American FactFinder page on the Census Bureau's Web site at <http://factfinder.census.gov> and under the heading **Getting Detailed Data** select **get data** for the American Community Survey (see Figure 5).

This gets you to the ACS data sets page where you would first select your data set. For this example, we want the **2006 American Community Survey**. As shown in Figure 6, once you select a data set, a list of data products appears. For this exercise, we would select **Detailed Tables**.

Figure 5. **Example of American FactFinder Web Page**

Source: U.S. Census Bureau, accessed at <http://www.census.gov/acs/www/Products/PUMS/>.

Figure 6. **Example of ACS Data Sets Web Page**

Source: U.S. Census Bureau, accessed at <http://factfinder.census.gov>.
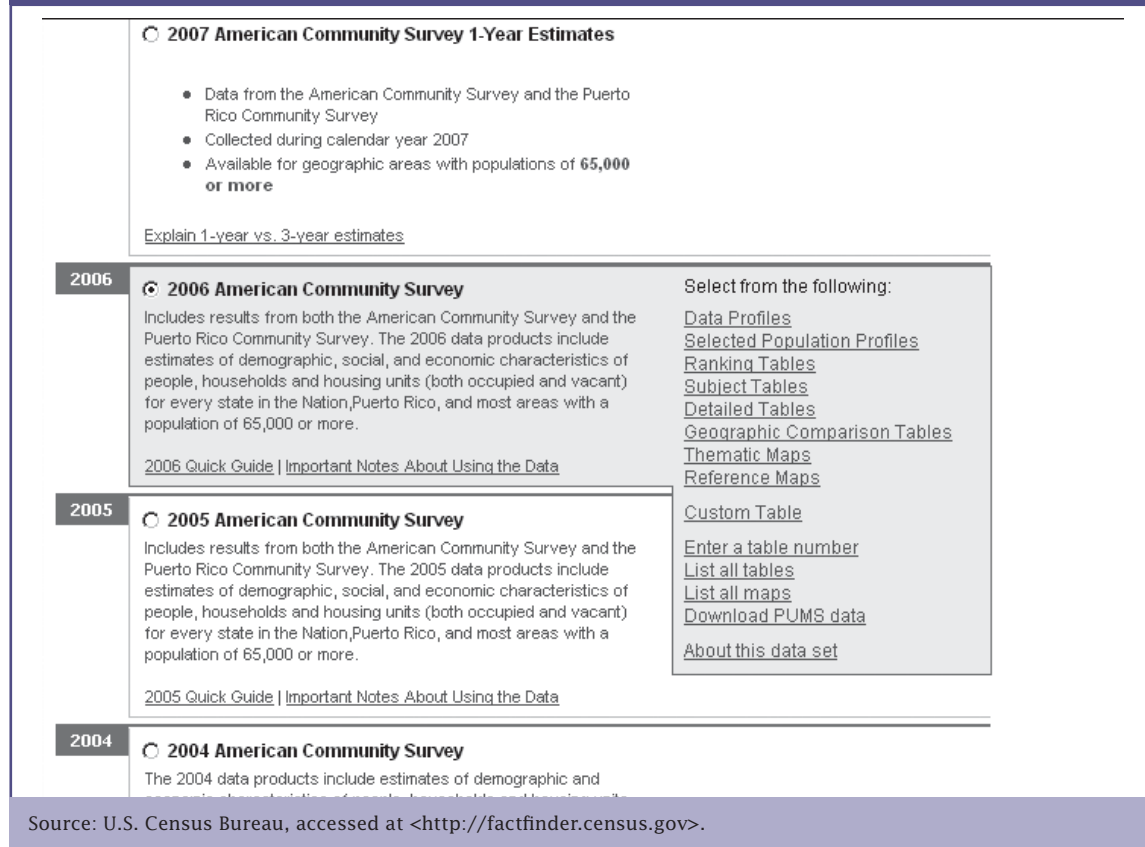
Now you must specify the geographic area or areas of interest. We are interested in state-level data so on the **Select Geography** page:

- **For geographic type** select **State**.
- For geographic areas select **All States**, **Add**, and then **Next**.

You are then requested to identify the specific table or tables. On the **Select Tables** page you will see a list of tables. See Figure 7 for an example. You can scroll down the full list or use the keyword or subject tabs to search for specific tables. For this exercise, entering "poverty" in the keyword box provides a shorter list of tables that include statistics on poverty. From this list:

- Select table **B17002. Ratio of Income to Poverty Level in the Past 12 Months**.
- Select **Add**.
- You can select more than one table from this list and add them to your request. Sometimes if you are unsure of which estimate you need, it makes sense to list a few tables.
- Click **Show Result**.

The **Detailed Tables** page displays the results of your query. A portion of this table is shown in Figure 8.

- The table is arranged with rows for the population by ratio of income to poverty level. The first row of data is for estimates of the Total Population for Whom Poverty Status Is Determined.

- The columns are organized by the selected geographic area or areas and contain point estimates and the margins of error for those estimates. There are a total of 52 geographic areas in the table: the 50 states; Washington, DC; and Puerto Rico.

- By selecting to download the output (see circled tab in Figure 8) you have a number of formatting options that are shown in Figure 9. For large queries involving multiple tables and many geographic areas, an alternative is to use the ACS Summary File's FTP server to download selected Summary Files at <http://www.census.gov/acs/www/Special/acsftp.html>. The FTP site contains documentation, data files, SAS files, and an example illustrating how to access and merge the relevant files for one detailed table for all geographic areas covered in the 2006 ACS.

Figure 7. **Example of Select Tables Page**

Source: U.S. Census Bureau, accessed at <http://factfinder.census.gov>.

Figure 8. **Example of ACS Detailed Tables Page**

Source: U.S. Census Bureau, accessed at <http://factfinder.census.gov>.

Figure 9. **Example of Download Web Page**

Detailed Tables ▶ Download

■ Select download options and click 'OK'   ⓘ Explain my choices

Download file format (data rows plus headings and footnotes)
- ⦿ Comma delimited (.csv)
  (.csv is compatible with spreadsheet programs such as Microsoft Excel)
- ○ Tab delimited (.lst)
- ○ Rich Text Format (.rtf)
- ◉ Microsoft Excel (.xls)

Options (only applies to presentation formats)
- ☐ Only the tables and geographies on the screen

Database compatible (data rows only) - the download file is
a **zip file** containing one or more data files and a geographic content file.
Geographic codes are included in the database compatible download.
- ○ Microsoft Excel (.xls)
- ○ Comma delimited (.txt)
- ○ Pipe delimited (.txt)

Options (only applies to database compatible formats)
- ☑ Include descriptive data element names

[OK]   [Cancel]   ⓘ Using FactFinder with Windows XP Service Pack 2

Source: U.S. Census Bureau, accessed at <http://factfinder.census.gov>.

Exercise 2: Access, Tabulate, and Download ACS Microdata Records

Task: Replicate 2007 ACS User Verification Tabulations

Data File: 2007 American Community Survey PUMS

Geographic Area: United States, States, and Puerto Rico

Background

DataFerrett is a software application for data extraction and mining. It is a joint effort developed and maintained by the Census Bureau and the Centers for Disease Control with major support from the Bureau of Labor Statistics and the National Center for Health Statistics. Information about DataFerrett, TheDataWeb, and downloads of the software are available from the Census Bureau at <http://dataferrett.census.gov/>. Some background on DataFerrett is assumed for the following exercise. For instruction on the basics of DataFerrett, refer to the tutorial on the DataFerrett Web page or the PUMS handbook in the *ACS Compass Products*.

This exercise uses DataFerrett to extract data from the 2007 ACS PUMS, tabulate selected variables from the population and housing records, and download the data to a desktop statistical software package. The task is to replicate the data in the **2007 ACS PUMS**

Estimates for User Verification. You can access this file from the ACS main page—access data tab by clicking on PUMS Files and PUMS Estimates for User Verification. As described earlier, these files are designed to help data users verify that they are processing PUMS data correctly. The User Verification file is presented in three formats: SAS Data Set, .lst text format, and .csv format. You need to download and display one of these formats. The file contains tabulations of people in housing units and group quarters, by sex and by age; and for housing units by occupancy status and tenure for the United States, 50 states, the District of Columbia, and Puerto Rico.

For this example, we will focus on three population variables from the 2007 ACS PUMS—age, sex, and relationship. This exercise illustrates a basic methodology that can be applied to all ACS PUMS variables. We want to use DataFerrett to produce estimates of the total population living in housing units and in institutional and noninstitutional group quarters (GQs). Specifically, we want the following tallies:

- Total population
- Total males and total females
- People aged 0–4, 5–9, 10–14, 15–19, 20–24, 25–34, 35–44, 45–54, 55–59, 60–64, 65–74, 75–84, 85 and older

Access DataFerrett

There are several ways to access DataFerrett. Links exist from the ACS PUMS page and from the American FactFinder page. From the American FactFinder page, select **TOOLS AND REFERENCES** from the left sidebar. Clicking on **Tools** brings you to a new page where, under **Data Extraction Tools** you can select **Census Bureau's DataFerrett**. Download DataFerrett (if necessary) and then log in. Click the **Get Data Now** box to begin creating your tabulation.

Select Data Set and Variables

You will start in DataFerrett under the tab **Step 1: Select DataSet & Variable**. The first step is to select your data set from the list provided. For this exercise, double click on **American Community Survey and Public Use Microdata Sample**. Then select **2007** and **View Variables**. The View Variables screen identifies the types of variables available for the selected data set. See Figure 10. If you need subnat-ional estimates you would need to check the box for **Selectable Geographies**. Similarly, if you need housing variables you would check that box. For our exercise, you only need population variables, so you only need to check that box. When you click on the magnifying glass at the bottom for **Search Variables**, a list of population variables is provided. You can scroll down the list to choose your variables or use the search tool at the top of the page to isolate a specific variable. If you are selecting more than one variable from the list, use Ctrl and Enter. For our example, select three variables: relationship (REL), sex (SEX), and age (AGEP).

Click on the **Browse/Select Highlighted Variables** box. A pop-up window will open to display details about these variables. Highlighting each variable, you can select all values by clicking **select**, **OK,** and **OK**. This moves the variables into your newly created Data Basket. There are applications when you might only need data for certain values—this is where you could identify the subset of interest. For this exercise we will need all values. After all selections are made, it's wise to save your work. Using the **File** option in the tool bar, save this file.

Recoding Variables

Clicking on the tab titled, **Step 2—Data Basket /Download/Make a Table**, brings you to a new screen that summarizes the variables in your Data Basket. These variables include detailed categories that we need to combine. This is called recoding. We want

Figure 10. **Example of DataFerrett View Variables Screen**

to recode the REL and AGEP variables into the specific categories in the verification file.

Starting with the REL variable, highlight **REL** and click on **Recode Variable(s)** in the right tool bar. Change the default of Recode1 to RecodeREL. The variable name will remain Recode1 but this new label will help you remember which specific recoded variable it represents. REL is a categorical variable with 14 values. Highlight categories 1–12 using the Ctrl key to identify all values representing people living in housing units. Click on the **Recode** button to set them all to a value of 1. Double-clicking on the label box allows you to relabel this new category as "Living in housing units." Similarly, recode Institutional and noninstitutional GQ population as 2 and 3 and relabel. Click **OK** to accept the recodes.

Recoding the age variable involves a slightly different approach since we want 5-year intervals from 1 to 24, 10-year intervals from 25 to 84, and a separate category for people 85 and older. Highlight **AGEP** and click **Recode Variable(s)**. Change the default label of Recode2 to RecodeAGEP. In the continuous variables box, the first line states **Within the range of 0 through** and the box is set to 99. Change 99 to 24. Then on the line that states **subgroups repeat by** set the box to 5. This recodes people ages 0 to 24 into 5-year age categories. Now the line states **Within the range of 25 through 99**. Use the same method to create 10-year intervals between 25 and 84 and then a final category for people 85 and older.

Creating a Table

Click on the **Make a Table** icon. The variables you have selected and the recode variables you have created are displayed. Click and drag the variables Recode1 (RecodeREL), SEX, and Recode2 (RecodeAGEP) to Column 1 in the order they are presented in the **Verify Tabulations** table. Be careful not to overlap the variables in column 1 or DataFerrett will think you want to cross tabulate them. Click on the green box— **GO Get Data**. Note that DataFerrett used the variable PWGT (Population Weight) to estimate the population values since the variables are all for population records. Housing unit records are weighted by the variable for housing unit weight. Tables for population and housing unit records need to be created separately in DataFerrett. The data provided are national totals. To produce data for additional geographies you would have needed to identify an additional geography variable.

Download the data set into a desktop statistical software package of your choice. Click the **Download** button to get an extract. (The Download button is located near the top of the page and has a picture of a ferret sitting on a hard drive.) A Download Data window pops up. Check the box next to Download Data. Choose

from among the six file formats offered. Actually, the formats will all produce ascii text files for the data records, but will also produce program statements in SAS, SPSS, or STATA, enabling you to read in the data records and create a system file. Depending on the size of the file you created, you can choose to compress or not and to process batch mode or not.

Create the rest of the table values for the housing unit characteristics. If you match the Verify Tabulations values, then you did it correctly.

Exercise 3: Compare ACS Poverty Data With Census 2000

Task: Compare Poverty Data in the ACS and Census

Data File: 2006 American Community Survey and 2000 Census Summary File 3

Geographic Area: Jefferson County, KY

Background

This exercise compares estimates of the number of school age children (5 to 17 years old) below the poverty level for Jefferson County, Kentucky, from the 2006 ACS with those from the Census 2000 Summary File 3 (SF3). It covers how to access comparable data on estimates of the number of children and how to determine whether the differences are statistically significant. A basic familiarity with using the American FactFinder interface to census data is assumed.

Access the Data

Go to the Census Bureau's ACS Web page. Select the **How to Use the Data** tab. From the left sidebar select **Comparing ACS Data to Other Sources** at <http://www.census.gov/acs/www/UseData /compACS.htm>. This page provides important information about comparing ACS data for a single year to ACS data for another year. It also provides guidance on comparing ACS data with data from Census 2000. Both general and item-specific guidance are provided. Scroll down to **Subject Area/Item 17 – Poverty Status**, and the row for "Poverty Status of All People in the Poverty Universe." Note that comparisons between the 2006 ACS and Census 2000 are identified as "compare with caution." The data are not directly comparable due to differences in reference period, data collection period, and residence rules—all of which have been discussed in this handbook.

Click on the link to **table crosswalk**. The first pair of tables is:

- Census 2000— P87: POVERTY STATUS IN 1999 BY AGE
- 2006 ACS—B17001: POVERTY STATUS IN THE PAST 12 MONTHS BY SEX BY AGE

Go to the American FactFinder page on the Census Bureau's Web site at <http://factfinder.census.gov> and under the heading **Getting Detailed Data** select **get data** for the Decennial Census. Select the button for **Census 2000 Summary File 3 (SF 3)—Sample Data**. To the right, a number of options appear under the heading **Select from the following**. Select **Enter a Table Number** and in the dialog box that appears enter P87 and click **GO**. For the geographic area, make the choice to select Jefferson County, Kentucky, and then click **GO**. Note that Table P87. POVERTY STATUS IN 1999 BY AGE [17] – Universe: Population for who poverty status is determined is displayed. Download and save this table.

Use the American FactFinder to select **2006 American Community Survey** and **Detailed Tables**. Depending on how you accessed the 2006 ACS, the geography selection you made could be saved or you will need to work through the selection options again. Using the table selection method **Show all tables**, scroll down to table B17001. Select and show the result. Note that table B17001. POVERTY STATUS IN THE PAST 12 MONTHS BY SEX BY AGE – Universe: Population for whom poverty status is determined is displayed. Download and save this table.

Conduct Statistical Testing

You now have two estimates of the number of school age children (5 to 17 years old) below the poverty level for Jefferson County, Kentucky. Before concluding if these two estimates are different you must consider the sampling errors associated with each estimate and determine if the difference is statistically significant. Appendix 4 describes the correct procedure for determining if the difference between two estimates is statistically significant. Algebraically, the significance test is expressed as:

$$\text{If } \left| \frac{\hat{X}_1 - \hat{X}_2}{\sqrt{SE_1^2 + SE_2^2}} \right| > Z_{CL},$$

then the difference between estimates $\hat{X}_1$ and $\hat{X}_2$ is statistically significant at the specified confidence level, CL

where $\hat{X}_i$ is estimate $i(=1,2)$

SE_i is the SE for the estimate $i(=1,2)$

Z_{CL} is the critical value for the desired confidence level (=1.645 for 90 percent, 1.960 for 95 percent, 2.576 for 99 percent).

In this exercise $\hat{X}_1$= 20,583 (Census 2000) and $\hat{X}_2$= 11,464 (2006 ACS). If Z has an absolute value greater than 1.645 then the estimates are statistically different at the 90-percent confidence level.

First determine the standard error (SE) of the estimate from the 2006 ACS. Appendix 3 details the steps needed to obtain Standard Errors for Derived Estimates. Algebraically, the MOE for the aggregated count is calculated as:

$$MOE_{agg} = \pm \sqrt{\sum_c MOE_c^2}$$

where MOE_c is the MOE of the cth component estimate.

Table B17001 (2006 ACS) presents the period estimates and their margins of error (MOE). Use the formula above to calculate the MOE of the derived poverty estimate for children aged 5 to 17 years old. In order to derive the SE, divide the MOE by 1.645. (Answer: the SE of the derived estimate of 11,464 is 1,205.)

Next you need to determine the SE of the estimate from the Census 2000 Summary File 3 (SF3). The publication TECHNICAL DOCUMENTATION: Census 2000 Summary File 3 Technical Documentation prepared by the Census Bureau, 2002 is available on the Census Bureau's Web site at <http://www.census.gov/prod/cen2000/doc /sf3.pdf>. Beginning on page 8-6 in **Chapter 8. Accuracy of the Data** is the section **Calculation of Standard Errors**. Follow the steps outlined in the section to determine the:

- Unadjusted SE computed using the formula below Table A on page 8-22 (Answer: 316).
- Observed sampling rate from Table P4 of SF3 (Answer: 17.1 percent).
- Design factor (Answer: 1.3—Poverty Status of Persons).
- SE (Answer: SE of the derived estimate of 20,583 is 411).

Referring back to the significance test, solve for Z. (Answer: 7.2) Since Z is greater than 1.645 we conclude that the difference in estimates of children aged 5 to 17 years who are below the poverty level in Jefferson County, Kentucky, between the Census 2000 and the ACS 2006 is statistically significant.

A Resource for Academic Researchers

The ACS offers a number of advantages to the research community. Data from the ACS are produced annually. The samples are far larger than those of the Current Population Survey (CPS) and the American Housing Survey (AHS), which were among the limited options for frequently updated annual data prior to the advent of the ACS. The data are comparable generally between years of the ACS and with caution for many items from Census 2000. The process of improving the ACS is ongoing, and the Census Bureau conducts methodological research of interest to academic researchers. A series of reports, papers, and presentations is publicly available on the Census Bureau's Web site. The Evaluation Report Series for the ACS can be accessed at <http://www.census.gov/acs/www/AdvMeth/Reports.htm>. A series of working papers and presentations is available at <http://www.census.gov/acs/www/AdvMeth/Papers/Papers1.htm>.

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).

American Community Survey Alert. This periodic electronic newsletter informs data users and other interested parties about news, events, data releases, congressional actions, and other developments associated with the ACS. See <http://www.census.gov/acs/www/Special/Alerts/Latest.htm>.

American FactFinder (AFF). An electronic system for access to and dissemination of Census Bureau data on the Internet. AFF offers prepackaged data products and user-selected data tables and maps from Census 2000, the 1990 Census of Population and Housing, the 1997 and 2002 Economic Censuses, the Population Estimates Program, annual economic surveys, and the ACS.

Block group. A subdivision of a census tract (or, prior to 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, including their structure, designations, and relationships to one another. See <http://www.census.gov/geo/www/index.html>.

Census tract. A small, relatively permanent statistical subdivision of a county delineated by a local committee of census data users for the purpose of presenting data. Census tract boundaries normally follow visible features, but may follow governmental unit boundaries and other nonvisible features; they always nest within counties. Designed to be relatively homogeneous units with respect to population characteristics, economic status, and living conditions at the time of establishment, census tracts average about 4,000 inhabitants.

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 profile. Comparison profiles are available from the American Community Survey for 1-year estimates beginning in 2007. These tables are available for the United States, the 50 states, the District of Columbia, and geographic areas with a population of more than 65,000.

Confidence interval. The sample estimate and its standard error permit the construction of a confidence interval that represents the degree of uncertainty about the estimate. A 90-percent confidence interval can be interpreted roughly as providing 90 percent certainty that the interval defined by the upper and lower bounds contains the true value of the characteristic.

Confidentiality. The guarantee made by law (Title 13, U.S. Code) to individuals who provide census information, regarding nondisclosure of that information to others.

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.

Controlled. During the ACS weighting process, the intercensal population and housing estimates are used as survey controls. Weights are adjusted so that ACS estimates conform to these controls.

Current Population Survey (CPS). The CPS is a monthly survey of about 50,000 households conducted by the Census Bureau for the Bureau of Labor Statistics. The CPS is the primary source of information on the labor force characteristics of the U.S. population.

Current residence. The concept used in the ACS to determine who should be considered a resident of a sample address. 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 geography. These products are available on the American FactFinder and the ACS Web site.

For users with data needs not met through the general purpose products, the Census Bureau offers "custom" tabulations on a cost-reimbursable basis, with the American Community Survey Custom Tabulation program. Custom tabulations 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 profiles. Detailed tables that provide summaries by social, economic, and housing characteristics. There is a new ACS demographic and housing units profile that should be used if official estimates from the Population Estimates Program are not available.

Detailed tables. Approximately 1,200 different tables that contain basic distributions of characteristics. These tables provide the most detailed data and are the basis for other ACS products.

Disclosure avoidance (DA). Statistical methods used in the tabulation of data prior to releasing data products to ensure the confidentiality of responses. See Confidentiality.

Estimates. Numerical values obtained from a statistical sample and assigned to a population parameter. Data produced from the ACS interviews are collected from samples of housing units. These data are used to produce estimates of the actual figures that would have been obtained by interviewing the entire population using the same methodology.

File Transfer Protocol (FTP) site. A Web site that allows data files to be downloaded from the Census Bureau Web site.

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

Geographic comparison tables. More than 80 single-variable tables comparing key indicators for geographies other than states.

Geographic summary level. A geographic summary level specifies the content and the hierarchical relationships of the geographic elements that are required to tabulate and summarize data. For example, the county summary level specifies the state-county hierarchy. Thus, both the state code and the county code are required to uniquely identify a county in the United States or Puerto Rico.

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 each other. The ACS collects data from people living in both housing units and GQ facilities.

Group quarters (GQ) population. The number of persons residing in GQ facilities.

Item allocation rates. Allocation is a method of imputation used when values for missing or inconsistent items cannot be derived from the existing response record. In these cases, the imputation must be based on other techniques such as using answers from other people in the household, other responding housing units, or people believed to have similar characteristics. Such donors are reflected in a table referred to as an allocation matrix. The rate is percentage of times this method is used.

Margin of error (MOE). Some ACS products provide an MOE instead of confidence intervals. An MOE is the difference between an estimate and its upper or lower confidence bounds. Confidence bounds can be created by adding the MOE to the estimate (for the upper bound) and subtracting the MOE from the estimate (for the lower bound). All published ACS MOE are based on a 90-percent confidence level.

Multiyear estimates. Three- and five-year estimates based on multiple years of ACS data. Three-year estimates will be published for geographic areas with a population of 20,000 or more. Five-year estimates will be published for all geographic areas down to the census block group level.

Narrative profile. A data product that includes easy-to-read descriptions for a particular geography.

Nonsampling error. Total survey error can be classified into two categories—sampling error and nonsampling error. Nonsampling error includes measurement errors due to interviewers, respondents, instruments, and mode; nonresponse error; coverage error; and processing error.

Period estimates. An estimate based on information collected over a period of time. For ACS the period is either 1 year, 3 years, or 5 years.

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

Population Estimates Program. Official Census Bureau estimates of the population of the United States, states, metropolitan areas, cities and towns, and counties; also official Census Bureau estimates of housing units (HUs).

Public Use Microdata Area (PUMA). An area that defines the extent of territory for which the Census Bureau releases Public Use Microdata Sample (PUMS) records.

Public Use Microdata Sample (PUMS) files. Computerized files that contain a sample of individual records, with identifying information removed, showing the population and housing characteristics of the units, and people included on those forms.

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 ACS data. The ACS releases four different quality measures 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. The ACS Quality Measures Web site provides these statistics each year. In addition, the coverage rates are also available for males and females separately.

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 series of rules that define who (if anyone) is considered to be a resident of a sample address for purposes of the survey or census.

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

Selected population profiles. An ACS data product that provides certain characteristics for a specific race or ethnic group (for example, Alaska Natives) or other population subgroup (for example, people aged 60 years and over). This data product is produced directly from the sample microdata (that is, not a derived product).

Single-year estimates. Estimates based on the set of ACS interviews conducted from January through December of a given calendar year. These estimates are published each year for geographic areas with a population of 65,000 or more.

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

Statistical significance. The determination of whether the difference between two estimates is not likely to be from random chance (sampling error) alone. This determination is based on both the estimates themselves and their standard errors. For ACS data, two estimates are "significantly different at the 90 percent level" if their difference is large enough to infer that there was a less than 10 percent chance that the difference came entirely from random variation.

Subject tables. Data products organized by subject area that present an overview of the information that analysts most often receive requests for from data users.

Summary files. Consist of detailed tables of Census 2000 social, economic, and housing characteristics compiled from a sample of approximately 19 million housing units (about 1 in 6 households) that received the Census 2000 long-form questionnaire.

Thematic maps. Display geographic variation in map format from the geographic ranking tables.

Three-year estimates. Estimates based on 3 years of ACS data. These estimates are meant to reflect the characteristics of a geographic area over the entire 3-year period. These estimates will be published for geographic areas with a population of 20,000 or more.

Appendix 1.

Understanding and Using ACS Single-Year and Multiyear Estimates

What Are Single-Year and Multiyear Estimates?

Understanding Period Estimates

The ACS produces period estimates of socioeconomic and housing characteristics. It is designed to provide estimates that describe the average characteristics of an area over a specific time period. In the case of ACS single-year estimates, the period is the calendar year (e.g., the 2007 ACS covers January through December 2007). In the case of ACS multiyear estimates, the period is either 3 or 5 calendar years (e.g., the 2005– 2007 ACS estimates cover January 2005 through December 2007, and the 2006–2010 ACS estimates cover January 2006 through December 2010). The ACS multiyear estimates are similar in many ways to the ACS single-year estimates, however they encompass a longer time period. As discussed later in this appendix, the differences in time periods between single-year and multiyear ACS estimates affect decisions about which set of estimates should be used for a particular analysis.

While one may think of these estimates as representing average characteristics over a single calendar year or multiple calendar years, it must be remembered that the 1-year estimates are not calculated as an average of 12 monthly values and the multiyear estimates are not calculated as the average of either 36 or 60 monthly values. Nor are the multiyear estimates calculated as the average of 3 or 5 single-year estimates. Rather, the ACS collects survey information continuously nearly every day of the year and then aggregates the results over a specific time period—1 year, 3 years, or 5 years. The data collection is spread evenly across the entire period represented so as not to over-represent any particular month or year within the period.

Because ACS estimates provide information about the characteristics of the population and housing for areas over an entire time frame, ACS single-year and multiyear estimates contrast with "point-in-time" estimates, such as those from the decennial census long-form samples or monthly employment estimates

from the Current Population Survey (CPS), which are designed to measure characteristics as of a certain date or narrow time period. For example, Census 2000 was designed to measure the characteristics of the population and housing in the United States based upon data collected around April 1, 2000, and thus its data reflect a narrower time frame than ACS data. The monthly CPS collects data for an even narrower time frame, the week containing the 12th of each month.

Implications of Period Estimates

Most areas have consistent population characteristics throughout the calendar year, and their period estimates may not look much different from estimates that would be obtained from a "point-in-time" survey design. However, some areas may experience changes in the estimated characteristics of the population, depending on when in the calendar year measurement occurred. For these areas, the ACS period estimates (even for a single-year) may noticeably differ from "point-in-time" estimates. The impact will be more noticeable in smaller areas where changes such as a factory closing can have a large impact on population characteristics, and in areas with a large physical event such as Hurricane Katrina's impact on the New Orleans area. This logic can be extended to better interpret 3-year and 5-year estimates where the periods involved are much longer. If, over the full period of time (for example, 36 months) there have been major or consistent changes in certain population or housing characteristics for an area, a period estimate for that area could differ markedly from estimates based on a "point-in-time" survey.

An extreme illustration of how the single-year estimate could differ from a "point-in-time" estimate within the year is provided in Table 1. Imagine a town on the Gulf of Mexico whose population is dominated by retirees in the winter months and by locals in the summer months. While the percentage of the population in the labor force across the entire year is about 45 percent (similar in concept to a period estimate), a "point-in-time" estimate for any particular month would yield estimates ranging from 20 percent to 60 percent.

Table 1. **Percent in Labor Force—Winter Village**											
Month											
Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.
20	20	40	60	60	60	60	60	60	50	30	20
Source: U.S. Census Bureau, Artificial Data.											

The important thing to keep in mind is that ACS single-year estimates describe the population and characteristics of an area for the full year, not for any specific day or period within the year, while ACS multiyear estimates describe the population and characteristics of an area for the full 3- or 5-year period, not for any specific day, period, or year within the multiyear time period.

Release of Single-Year and Multiyear Estimates

The Census Bureau has released single-year estimates from the full ACS sample beginning with data from the 2005 ACS. ACS 1-year estimates are published annually for geographic areas with populations of 65,000 or more. Beginning in 2008 and encompassing 2005–2007, the Census Bureau will publish annual ACS 3-year estimates for geographic areas with populations of 20,000 or more. Beginning in 2010, the Census Bureau will release ACS 5-year estimates

(encompassing 2005–2009) for all geographic areas —down to the tract and block group levels. While eventually all three data series will be available each year, the ACS must collect 5 years of sample before that final set of estimates can be released. This means that in 2008 only 1-year and 3-year estimates are available for use, which means that data are only available for areas with populations of 20,000 and greater.

New issues will arise when multiple sets of multiyear estimates are released. The multiyear estimates released in consecutive years consist mostly of overlapping years and shared data. As shown in Table 2, consecutive 3-year estimates contain 2 years of overlapping coverage (for example, the 2005–2007 ACS estimates share 2006 and 2007 sample data with the 2006–2008 ACS estimates) and consecutive 5-year estimates contain 4 years of overlapping coverage.

Table 2. **Sets of Sample Cases Used in Producing ACS Multiyear Estimates**					
Type of estimate	Year of Data Release				
	2008	2009	2010	2011	2012
	Years of Data Collection				
3-year estimates	2005–2007	2006–2008	2007–2009	2008–2010	2009–2011
5-year estimates	Not Available	Not Available	2005–2009	2006–2010	2007–2011
Source: U.S. Census Bureau.					

Differences Between Single-Year and Multiyear ACS Estimates

Currency

Single-year estimates provide more current information about areas that have changing population and/or housing characteristics because they are based on the most current data—data from the past year. In contrast, multiyear estimates provide less current information because they are based on both data from the previous year and data that are 2 and 3 years old. As noted earlier, for many areas with minimal change taking place, using the "less current" sample used to produce the multiyear estimates may not have a substantial influence on the estimates. However, in areas experiencing major changes over a given time period, the multiyear estimates may be quite different from the single-year estimates for any of the individual years. Single-year and multiyear estimates are not expected to be the same because they are based on data from two different time periods. This will be true even if the ACS

single year is the midyear of the ACS multiyear period (e.g., 2007 single year, 2006–2008 multiyear).

For example, suppose an area has a growing Hispanic population and is interested in measuring the percent of the population who speak Spanish at home. Table 3 shows a hypothetical set of 1-year and 3-year estimates. Comparing data by release year shows that for an area such as this with steady growth, the 3-year estimates for a period are seen to lag behind the estimates for the individual years.

Reliability

Multiyear estimates are based on larger sample sizes and will therefore be more reliable. The 3-year estimates are based on three times as many sample cases as the 1-year estimates. For some characteristics this increased sample is needed for the estimates to be reliable enough for use in certain applications. For other characteristics the increased sample may not be necessary.

Table 3. Example of Differences in Single- and Multiyear Estimates—Percent of Population Who Speak Spanish at Home

Year of data release	1-year estimates		3-year estimates	
	Time period	Estimate	Time period	Estimate
2003	2002	13.7	2000–2002	13.4
2004	2003	15.1	2001–2003	14.4
2005	2004	15.9	2002–2004	14.9
2006	2005	16.8	2003–2005	15.9

Source: U.S. Census Bureau, Artificial Data.

Multiyear estimates are the only type of estimates available for geographic areas with populations of less than 65,000. Users may think that they only need to use multiyear estimates when they are working with small areas, but this isn't the case. Estimates for large geographic areas benefit from the increased sample resulting in more precise estimates of population and housing characteristics, especially for subpopulations within those areas.

In addition, users may determine that they want to use single-year estimates, despite their reduced reliability, as building blocks to produce estimates for meaningful higher levels of geography. These aggregations will similarly benefit from the increased sample sizes and gain reliability.

Deciding Which ACS Estimate to Use

Three primary uses of ACS estimates are to understand the characteristics of the population of an area for local planning needs, make comparisons across areas, and assess change over time in an area. Local planning could include making local decisions such as where to locate schools or hospitals, determining the need for services or new businesses, and carrying out transportation or other infrastructure analysis. In the past, decennial census sample data provided the most comprehensive information. However, the currency of those data suffered through the intercensal period, and the ability to assess change over time was limited. ACS estimates greatly improve the currency of data for understanding the characteristics of housing and population and enhance the ability to assess change over time.

Several key factors can guide users trying to decide whether to use single-year or multiyear ACS estimates for areas where both are available: intended use of the estimates, precision of the estimates, and currency of

the estimates. All of these factors, along with an understanding of the differences between single-year and multiyear ACS estimates, should be taken into consideration when deciding which set of estimates to use.

Understanding Characteristics

For users interested in obtaining estimates for small geographic areas, multiyear ACS estimates will be the only option. For the very smallest of these areas (less than 20,000 population), the only option will be to use the 5-year ACS estimates. Users have a choice of two sets of multiyear estimates when analyzing data for small geographic areas with populations of at least 20,000. Both 3-year and 5-year ACS estimates will be available. Only the largest areas with populations of 65,000 and more receive all three data series.

The key trade-off to be made in deciding whether to use single-year or multiyear estimates is between currency and precision. In general, the single-year estimates are preferred, as they will be more relevant to the current conditions. However, the user must take into account the level of uncertainty present in the single-year estimates, which may be large for small subpopulation groups and rare characteristics. While single-year estimates offer more current estimates, they also have higher sampling variability. One measure, the coefficient of variation (CV) can help you determine the fitness for use of a single-year estimate in order to assess if you should opt instead to use the multiyear estimate (or if you should use a 5-year estimate rather than a 3-year estimate). The CV is calculated as the ratio of the standard error of the estimate to the estimate, times 100. A single-year estimate with a small CV is usually preferable to a multiyear estimate as it is more up to date. However, multiyear estimates are an alternative option when a single-year estimate has an unacceptably high CV.

Table 4 illustrates how to assess the reliability of 1-year estimates in order to determine if they should be used. The table shows the percentage of households where Spanish is spoken at home for ACS test counties Broward, Florida, and Lake, Illinois. The standard errors and CVs associated with those estimates are also shown.

In this illustration, the CV for the single-year estimate in Broward County is 1.0 percent (0.2/19.9) and in Lake County is 1.3 percent (0.2/15.9). Both are sufficiently small to allow use of the more current single-year estimates.

Single-year estimates for small subpopulations (e.g., families with a female householder, no husband, and related children less than 18 years) will typically have larger CVs. In general, multiyear estimates are preferable to single-year estimates when looking at estimates for small subpopulations.

For example, consider Sevier County, Tennessee, which had an estimated population of 76,632 in 2004 according to the Population Estimates Program. This population is larger than the Census Bureau's 65,000-population requirement for publishing 1-year estimates. However, many subpopulations within this geographic area will be much smaller than 65,000. Table 5 shows an estimated 21,881 families in Sevier County based on the 2000–2004 multiyear estimate; but only 1,883 families with a female householder, no

husband present, with related children under 18 years. Not surprisingly, the 2004 ACS estimate of the poverty rate (38.3 percent) for this subpopulation has a large standard error (SE) of 13.0 percentage points. Using this information we can determine that the CV is 33.9 percent (13.0/38.3).

For such small subpopulations, users obtain more precision using the 3-year or 5-year estimate. In this example, the 5-year estimate of 40.2 percent has an SE of 4.9 percentage points that yields a CV of 12.2 percent (4.9/40.2), and the 3-year estimate of 40.4 percent has an SE of 6.8 percentage points which yields a CV of 16.8 percent (6.8/40.4).

Users should think of the CV associated with an estimate as a way to assess "fitness for use." The CV threshold that an individual should use will vary based on the application. In practice there will be many estimates with CVs over desirable levels. A general guideline when working with ACS estimates is that, while data are available at low geographic levels, in situations where the CVs for these estimates are high, the reliability of the estimates will be improved by aggregating such estimates to a higher geographic level. Similarly, collapsing characteristic detail (for example, combining individual age categories into broader categories) can allow you to improve the reliability of the aggregate estimate, bringing the CVs to a more acceptable level.

Table 4. Example of How to Assess the Reliability of Estimates—Percent of Population Who Speak Spanish at Home

County	Estimate	Standard error	Coefficient of variation
Broward County, FL	19.9	0.2	1.0
Lake County, IL	15.9	0.2	1.3

Source: U.S. Census Bureau, Multiyear Estimates Study data.

Table 5. Percent in Poverty by Family Type for Sevier County, TN

	2000–2004	2000–2004		2002–2004		2004	
	Total family type	Pct. in poverty	SE	Pct. in poverty	SE	Pct. in poverty	SE
All families	21,881	9.5	0.8	9.7	1.3	10.0	2.3
With related children under 18 years	9,067	15.3	1.5	16.5	2.4	17.8	4.5
Married-couple families	17,320	5.8	0.7	5.4	0.9	7.9	2.0
With related children under 18 years	6,633	7.7	1.2	7.3	1.7	12.1	3.9
Families with female householder, no husband	3,433	27.2	3.0	26.7	4.8	19.0	7.2
With related children under 18 years	1,883	40.2	4.9	40.4	6.8	38.3	13.0

Source: U.S. Census Bureau, Multiyear Estimates Study data.

Making Comparisons

Often users want to compare the characteristics of one area to those of another area. These comparisons can be in the form of rankings or of specific pairs of comparisons. Whenever you want to make a comparison between two different geographic areas you need to take the type of estimate into account. It is important that comparisons be made within the same estimate type. That is, 1-year estimates should only be compared with other 1-year estimates, 3-year estimates should only be compared with other 3-year estimates, and 5-year estimates should only be compared with other 5-year estimates.

You certainly can compare characteristics for areas with populations of 30,000 to areas with populations of 100,000 but you should use the data set that they have in common. In this example you could use the 3-year or the 5-year estimates because they are available for areas of 30,000 and areas of 100,000.

Assessing Change

Users are encouraged to make comparisons between sequential single-year estimates. Specific guidance on making these comparisons and interpreting the results are provided in Appendix 4. Starting with the 2007 ACS, a new data product called the comparison profile will do much of the statistical work to identify statistically significant differences between the 2007 ACS and the 2006 ACS.

As noted earlier, caution is needed when using multiyear estimates for estimating year-to-year change in a particular characteristic. This is because roughly two-thirds of the data in a 3-year estimate overlap with the data in the next year's 3-year estimate (the overlap is roughly four-fifths for 5-year estimates). Thus, as shown in Figure 1, when comparing 2006–2008 3-year estimates with 2007–2009 3-year estimates, the differences in overlapping multiyear estimates are driven by differences in the nonoverlapping years. A data user interested in comparing 2009 with 2008 will not be able to isolate those differences using these two successive 3-year estimates. Figure 1 shows that the difference in these two estimates describes the difference between 2009 and 2006. While the interpretation of this difference is difficult, these comparisons can be made with caution. Users who are interested in comparing overlapping multiyear period estimates should refer to Appendix 4 for more information.

Figure 1. **Data Collection Periods for 3–Year Estimates**
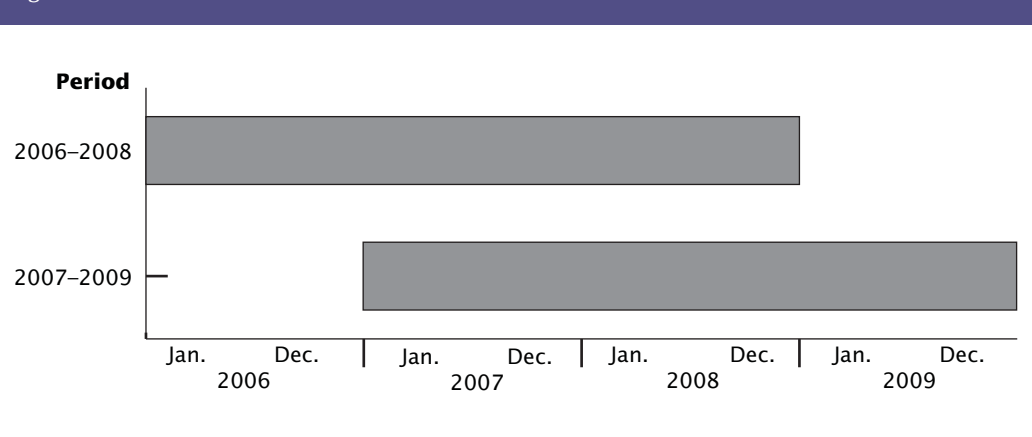
Source: U.S. Census Bureau.

Variability in single-year estimates for smaller areas (near the 65,000-publication threshold) and small subgroups within even large areas may limit the ability to examine trends. For example, single-year estimates for a characteristic with a high CV vary from year to year because of sampling variation obscuring an underlying trend. In this case, multiyear estimates may be useful for assessing an underlying, long-term trend. Here again, however, it must be recognized that because the multiyear estimates have an inherent smoothing, they will tend to mask rapidly developing changes. Plotting the multiyear estimates as representing the middle year is a useful tool to illustrate the smoothing effect of the multiyear weighting methodology. It also can be used to assess the "lagging effect" in the multiyear estimates. As a general rule, users should not consider a multiyear estimate as a proxy for the middle year of the period. However, this could be the case under some specific conditions, as is the case when an area is experiencing growth in a linear trend.

As Figure 2 shows, while the single-year estimates fluctuate from year to year without showing a smooth trend, the multiyear estimates, which incorporate data from multiple years, evidence a much smoother trend across time.

Figure 2. **Civilian Veterans, County X Single-Year, Multiyear Estimates**
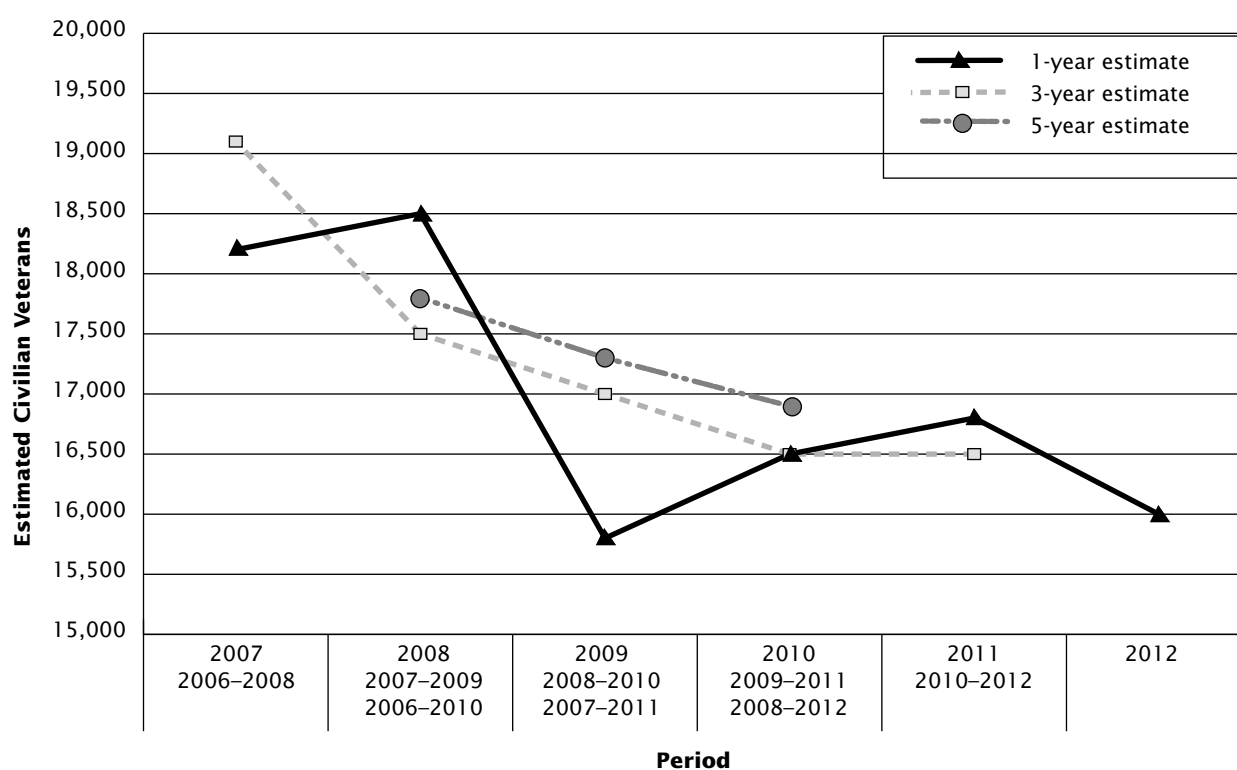
Source: U.S. Census Bureau. Based on data from the Multiyear Estimates Study.

Summary of Guidelines

Multiyear estimates should, in general, be used when single-year estimates have large CVs or when the precision of the estimates is more important than the currency of the data. Multiyear estimates should also be used when analyzing data for smaller geographies and smaller populations in larger geographies. Multiyear estimates are also of value when examining change over nonoverlapping time periods and for smoothing data trends over time.

Single-year estimates should, in general, be used for larger geographies and populations when currency is more important than the precision of the estimates. Single-year estimates should be used to examine year-to-year change for estimates with small CVs. Given the availability of a single-year estimate, calculating the CV provides useful information to determine if the single-year estimate should be used. For areas believed to be experiencing rapid changes in a characteristic, single-year estimates should generally be used rather than multiyear estimates as long as the CV for the single-year estimate is reasonable for the specific usage.

Local area variations may occur due to rapidly occurring changes. As discussed previously, multiyear estimates will tend to be insensitive to such changes when they first occur. Single-year estimates, if associ-

ated with sufficiently small CVs, can be very valuable in identifying and studying such phenomena. Graphing trends for such areas using single-year, 3-year, and 5-year estimates can take advantage of the strengths of each set of estimates while using other estimates to compensate for the limitations of each set.

Figure 3 provides an illustration of how the various ACS estimates could be graphed together to better understand local area variations.

The multiyear estimates provide a smoothing of the upward trend and likely provide a better portrayal of the change in proportion over time. Correspondingly, as the data used for single-year estimates will be used in the multiyear estimates, an observed change in the upward direction for consecutive single-year estimates could provide an early indicator of changes in the underlying trend that will be seen when the multiyear estimates encompassing the single years become available.

We hope that you will follow these guidelines to determine when to use single-year versus multiyear estimates, taking into account the intended use and CV associated with the estimate. The Census Bureau encourages you to include the MOE along with the estimate when producing reports, in order to provide the reader with information concerning the uncertainty associated with the estimate.

Figure 3. **Proportion of Population With Bachelor's Degree or Higher, City X Single-Year, Multiyear Estimates**
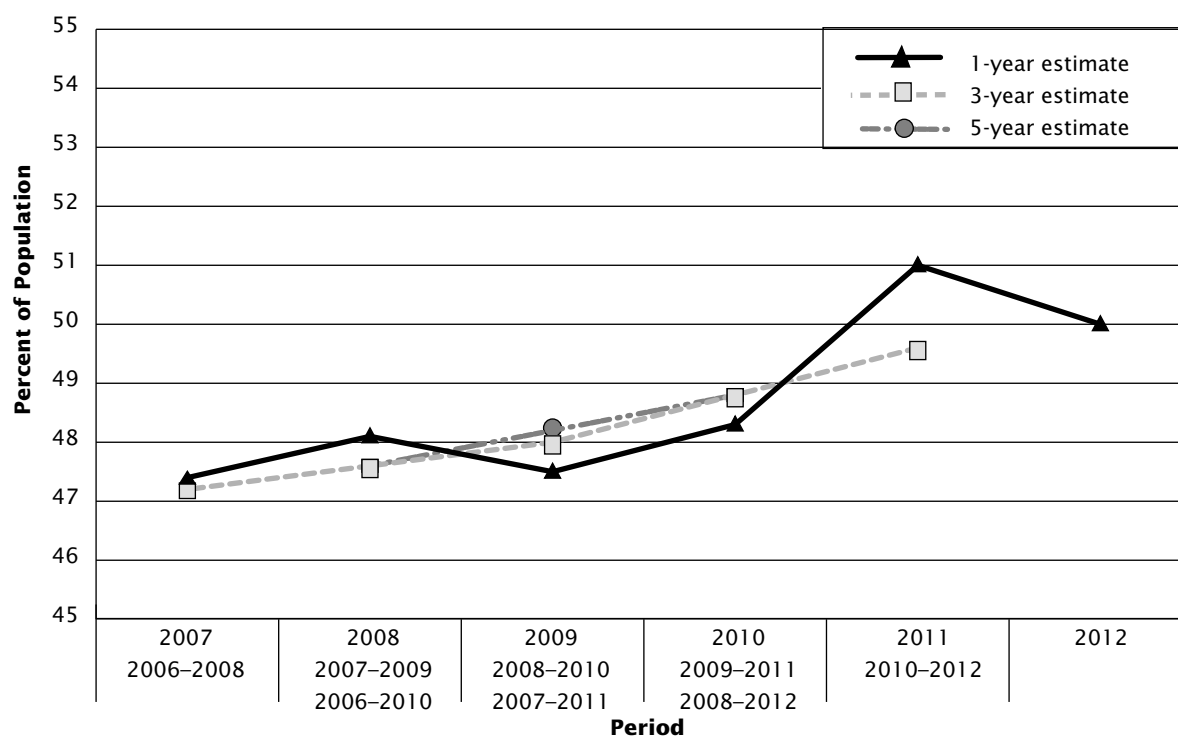
Source: U.S. Census Bureau. Based on data from the Multiyear Estimates Study.

Appendix 2.

Differences Between ACS and Decennial Census Sample Data

There are many similarities between the methods used in the decennial census sample and the ACS. Both the ACS and the decennial census sample data are based on information from a sample of the population. The data from the Census 2000 sample of about one-sixth of the population were collected using a "long-form" questionnaire, whose content was the model for the ACS. While some differences exist in the specific Census 2000 question wording and that of the ACS, most questions are identical or nearly identical. Differences in the design and implementation of the two surveys are noted below with references provided to a series of evaluation studies that assess the degree to which these differences are likely to impact the estimates. As noted in Appendix 1, the ACS produces period estimates and these estimates do not measure characteristics for the same time frame as the decennial census estimates, which are interpreted to be a snapshot of April 1 of the census year. Additional differences are described below.

Residence Rules, Reference Periods, and Definitions

The fundamentally different purposes of the ACS and the census, and their timing, led to important differences in the choice of data collection methods. For example, the residence rules for a census or survey determine the sample unit's occupancy status and household membership. Defining the rules in a dissimilar way can affect those two very important estimates. The Census 2000 residence rules, which determined where people should be counted, were based on the principle of "usual residence" on April 1, 2000, in keeping with the focus of the census on the requirements of congressional apportionment and state redistricting. To accomplish this the decennial census attempts to restrict and determine a principal place of residence on one specific date for everyone enumerated. The ACS residence rules are based on a "current residence" concept since data are collected continuously throughout the entire year with responses provided relative to the continuously changing survey interview dates. This method is consistent with the goal that the ACS produce estimates that reflect annual averages of the characteristics of all areas.

Estimates produced by the ACS are not measuring exactly what decennial samples have been measuring. The ACS yearly samples, spread over 12 months, collect information that is anchored to the day on which the sampled unit was interviewed, whether it is the day that a mail questionnaire is completed or the day that an interview is conducted by telephone or personal visit. Individual questions with time references such as

"last week" or "the last 12 months" all begin the reference period as of this interview date. Even the information on types and amounts of income refers to the 12 months prior to the day the question is answered. ACS interviews are conducted just about every day of the year, and all of the estimates that the survey releases are considered to be averages for a specific time period. The 1-year estimates reflect the full calendar year; 3-year and 5-year estimates reflect the full 36- or 60-month period.

Most decennial census sample estimates are anchored in this same way to the date of enumeration. The most obvious difference between the ACS and the census is the overall time frame in which they are conducted. The census enumeration time period is less than half the time period used to collect data for each single-year ACS estimate. But a more important difference is that the distribution of census enumeration dates are highly clustered in March and April (when most census mail returns were received) with additional, smaller clusters seen in May and June (when nonresponse follow-up activities took place).

This means that the data from the decennial census tend to describe the characteristics of the population and housing in the March through June time period (with an overrepresentation of March/April) while the ACS characteristics describe the characteristics nearly every day over the full calendar year.

Census Bureau analysts have compared sample estimates from Census 2000 with 1-year ACS estimates based on data collected in 2000 and 3-year ACS estimates based on data collected in 1999–2001 in selected counties. A series of reports summarize their findings and can be found at <http://www.census.gov/acs/www/AdvMeth/Reports.htm>. In general, ACS estimates were found to be quite similar to those produced from decennial census data.

More on Residence Rules

Residence rules determine which individuals are considered to be residents of a particular housing unit or group quarters. While many people have definite ties to a single housing unit or group quarters, some people may stay in different places for significant periods of time over the course of the year. For example, migrant workers move with crop seasons and do not live in any one location for the entire year. Differences in treatment of these populations in the census and ACS can lead to differences in estimates of the characteristics of some areas.

For the past several censuses, decennial census residence rules were designed to produce an accurate

count of the population as of Census Day, April 1, while the ACS residence rules were designed to collect representative information to produce annual average estimates of the characteristics of all kinds of areas. When interviewing the population living in housing units, the decennial census uses a "usual residence" rule to enumerate people at the place where they live or stay most of the time as of April 1. 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. The residence rules governing the census enumerations of people in group quarters depend on the type of group quarter and where permitted, whether people claim a "usual residence" elsewhere. The ACS applies a straight de facto residence rule to every type of group quarter. Everyone living or staying in a group quarter on the day it is visited by an ACS interviewer is eligible to be sampled and interviewed for the survey. Further information on residence rules can be found at <http://www.census.gov/acs/www/AdvMeth/CollProc/CollProc1.htm>.

The differences in the ACS and census data as a consequence of the different residence rules are most likely minimal for most areas and most characteristics. However, for certain segments of the population the usual and current residence concepts could result in different residence decisions. Appreciable differences may occur in areas where large proportions of the total population spend several months of the year in what would not be considered their residence under decennial census rules. In particular, data for areas that include large beach, lake, or mountain vacation areas may differ appreciably between the census and the ACS if populations live there for more than 2 months.

More on Reference Periods

The decennial census centers its count and its age distributions on a reference date of April 1, the assumption being that the remaining basic demographic questions also reflect that date, regardless of whether the enumeration is conducted by mail in March or by a field follow-up in July. However, nearly all questions are anchored to the date the interview is provided. Questions with their own reference periods, such as "last week," are referring to the week prior to the interview date. The idea that all census data reflect the characteristics as of April 1 is a myth. Decennial census samples actually provide estimates based on aggregated data reflecting the entire period of decennial data collection, and are greatly influenced by delivery dates of mail questionnaires, success of mail response, and data collection schedules for nonresponse follow-up. The ACS reference periods are, in many ways, similar to those in the census in that they reflect the circumstances on the day the data are collected and the individual reference periods of questions relative to that date. However, the ACS estimates represent the average characteristics over a full year (or sets of years), a different time, and reference period than the census.

Some specific differences in reference periods between the ACS and the decennial census are described below. Users should consider the potential impact these different reference periods could have on distributions when comparing ACS estimates with Census 2000.

Those who are interested in more information about differences in reference periods should refer to the Census Bureau's guidance on comparisons that contrasts for each question the specific reference periods used in Census 2000 with those used in the ACS. See <http://www.census.gov/acs/www/UseData/compACS.htm>.

Income Data

To estimate annual income, the Census 2000 long-form sample used the calendar year prior to Census Day as the reference period, and the ACS uses the 12 months prior to the interview date as the reference period. Thus, while Census 2000 collected income information for calendar year 1999, the ACS collects income information for the 12 months preceding the interview date. The responses are a mixture of 12 reference periods ranging from, in the case of the 2006 ACS single-year estimates, the full calendar year 2005 through November 2006. The ACS income responses for each of these reference periods are individually inflation-adjusted to represent dollar values for the ACS collection year.

School Enrollment

The school enrollment question on the ACS asks if a person had "at any time in the last 3 months attended a school or college." A consistent 3-month reference period is used for all interviews. In contrast, Census 2000 asked if a person had "at any time since February 1 attended a school or college." Since Census 2000 data were collected from mid-March to late-August, the reference period could have been as short as about 6 weeks or as long as 7 months.

Utility Costs

The reference periods for two utility cost questions—gas and electricity—differ between Census 2000 and the ACS. The census asked for annual costs, while the ACS asks for the utility costs in the previous month.

Definitions

Some data items were collected by both the ACS and the Census 2000 long form with slightly different definitions that could affect the comparability of the estimates for these items. One example is annual costs for a mobile home. Census 2000 included installment loan costs in

the total annual costs but the ACS does not. In this example, the ACS could be expected to yield smaller estimates than Census 2000.

Implementation

While differences discussed above were a part of the census and survey design objectives, other differences observed between ACS and census results were not by design, but due to nonsampling error—differences related to how well the surveys were conducted. Appendix 6 explains nonsampling error in more detail.

The ACS and the census experience different levels and types of coverage error, different levels and treatment of unit and item nonresponse, and different instances of measurement and processing error. Both Census 2000 and the ACS had similar high levels of survey coverage and low levels of unit nonresponse. Higher levels of unit nonresponse were found in the nonresponse follow-up stage of Census 2000. Higher item nonresponse rates were also found in Census 2000. Please see <http://www.census.gov/acs/www/AdvMeth/Reports.htm> for detailed comparisons of these measures of survey quality.

Appendix 3.

Measures of Sampling Error

All survey and census estimates include some amount of error. Estimates generated from sample survey data have uncertainty associated with them due to their being based on a sample of the population rather than the full population. This uncertainty, referred to as sampling error, means that the estimates derived from a sample survey will likely differ from the values that would have been obtained if the entire population had been included in the survey, as well as from values that would have been obtained had a different set of sample units been selected. All other forms of error are called nonsampling error and are discussed in greater detail in Appendix 6.

Sampling error can be expressed quantitatively in various ways, four of which are presented in this appendix—standard error, margin of error, confidence interval, and coefficient of variation. As the ACS estimates are based on a sample survey of the U.S. population, information about the sampling error associated with the estimates must be taken into account when analyzing individual estimates or comparing pairs of estimates across areas, population subgroups, or time periods. The information in this appendix describes each of these sampling error measures, explaining how they differ and how each should be used. It is intended to assist the user with analysis and interpretation of ACS estimates. Also included are instructions on how to compute margins of error for user-derived estimates.

Sampling Error Measures and Their Derivations

Standard Errors

A standard error (SE) measures the variability of an estimate due to sampling. Estimates derived from a sample (such as estimates from the ACS or the decennial census long form) will generally not equal the population value, as not all members of the population were measured in the survey. The SE provides a quantitative measure of the extent to which an estimate derived from the sample survey can be expected to deviate from this population value. It is the foundational measure from which other sampling error measures are derived. The SE is also used when comparing estimates to determine whether the differences between the estimates can be said to be statistically significant.

A very basic example of the standard error is a population of three units, with values of 1, 2, and 3. The average value for this population is 2. If a simple random sample of size two were selected from this population, the estimates of the average value would be 1.5 (units with values of 1 and 2 selected), 2 (units with values

of 1 and 3 selected), or 2.5 (units with values of 2 and 3 selected). In this simple example, two of the three samples yield estimates that do not equal the population value (although the average of the estimates across all possible samples do equal the population value). The standard error would provide an indication of the extent of this variation.

The SE for an estimate depends upon the underlying variability in the population for the characteristic and the sample size used for the survey. In general, the larger the sample size, the smaller the standard error of the estimates produced from the sample. This relationship between sample size and SE is the reason ACS estimates for less populous areas are only published using multiple years of data: to take advantage of the larger sample size that results from aggregating data from more than one year.

Margins of Error

A margin of error (MOE) describes the precision of the estimate at a given level of confidence. The confidence level associated with the MOE indicates the likelihood that the sample estimate is within a certain distance (the MOE) from the population value. Confidence levels of 90 percent, 95 percent, and 99 percent are commonly used in practice to lessen the risk associated with an incorrect inference. The MOE provides a concise measure of the precision of the sample estimate in a table and is easily used to construct confidence intervals and test for statistical significance.

The Census Bureau statistical standard for published data is to use a 90-percent confidence level. Thus, the MOEs published with the ACS estimates correspond to a 90-percent confidence level. However, users may want to use other confidence levels, such as 95 percent or 99 percent. The choice of confidence level is usually a matter of preference, balancing risk for the specific application, as a 90-percent confidence level implies a 10 percent chance of an incorrect inference, in contrast with a 1 percent chance if using a 99-percent confidence level. Thus, if the impact of an incorrect conclusion is substantial, the user should consider increasing the confidence level.

One commonly experienced situation where use of a 95 percent or 99 percent MOE would be preferred is when conducting a number of tests to find differences between sample estimates. For example, if one were conducting comparisons between male and female incomes for each of 100 counties in a state, using a 90-percent confidence level would imply that 10 of the comparisons would be expected to be found significant even if no differences actually existed. Using a 99-percent confidence level would reduce the likelihood of this kind of false inference.

Calculating Margins of Error for Alternative Confidence Levels

If you want to use an MOE corresponding to a confidence level other than 90 percent, the published MOE can easily be converted by multiplying the published MOE by an adjustment factor. If the desired confidence level is 95 percent, then the factor is equal to 1.960/1.645.[1] If the desired confidence level is 99 percent, then the factor is equal to 2.576/1.645.

Conversion of the published ACS MOE to the MOE for a different confidence level can be expressed as

$$MOE_{95} = \frac{1.960}{1.645} MOE_{ACS}$$

$$MOE_{99} = \frac{2.576}{1.645} MOE_{ACS}$$

where MOE_{ACS} is the ACS published 90 percent MOE for the estimate.

Factors Associated With Margins of Error for Commonly Used Confidence Levels
90 Percent: 1.645 95 Percent: 1.960 99 Percent: 2.576
Census Bureau standard for published MOE is 90 percent.

For example, the ACS published MOE for the 2006 ACS estimated number of civilian veterans in the state of Virginia is ±12,357. The MOE corresponding to a 95-percent confidence level would be derived as follows:

$$MOE_{95} = \frac{1.960}{1.645}(\pm 12,357) = \pm 14,723$$

Deriving the Standard Error From the MOE

When conducting exact tests of significance (as discussed in Appendix 4) or calculating the CV for an estimate, the SEs of the estimates are needed. To derive the SE, simply divide the positive value of the published MOE by 1.645.[2]

Derivation of SEs can thus be expressed as

$$SE = \frac{MOE_{ACS}}{1.645}$$

where MOE_{ACS} is the positive value of the ACS published MOE for the estimate.

For example, the ACS published MOE for estimated number of civilian veterans in the state of Virginia from the 2006 ACS is ±12,357. The SE for the estimate would be derived as

$$SE = \frac{12,357}{1.645} = 7,512$$

Confidence Intervals

A confidence interval (CI) is a range that is expected to contain the average value of the characteristic that would result over all possible samples with a known probability. This probability is called the "level of confidence" or "confidence level." CIs are useful when graphing estimates to display their sampling variabilites. The sample estimate and its MOE are used to construct the CI.

Constructing a Confidence Interval From a Margin of Error

To construct a CI at the 90-percent confidence level, the published MOE is used. The CI boundaries are determined by adding to and subtracting from a sample estimate, the estimate's MOE.

For example, if an estimate of 20,000 had an MOE at the 90-percent confidence level of ±1,645, the CI would range from 18,355 (20,000 − 1,645) to 21,645 (20,000 + 1,645).

For CIs at the 95-percent or 99-percent confidence level, the appropriate MOE must first be derived as explained previously.

Construction of the lower and upper bounds for the CI can be expressed as

$$L_{CL} = \hat{X} - MOE_{CL}$$
$$U_{CL} = \hat{X} + MOE_{CL}$$

where $\hat{X}$ is the ACS estimate and

MOE_{CL} is the positive value of the MOE for the estimate at the desired confidence level.

The CI can thus be expressed as the range

$$CI_{CL} = \left(L_{CL}, U_{CL}\right).[3]$$

---

[1] The value 1.65 must be used for ACS single-year estimates for 2005 or earlier, as that was the value used to derive the published margin of error from the standard error in those years.

[2] If working with ACS 1-year estimates for 2005 or earlier, use the value 1.65 rather than 1.645 in the adjustment factor.

[3] Users are cautioned to consider logical boundaries when creating confidence intervals from the margins of error. For example, a small population estimate may have a calculated lower bound less than zero. A negative number of persons doesn't make sense, so the lower bound should be set to zero instead.

For example, to construct a CI at the 95-percent confidence level for the number of civilian veterans in the state of Virginia in 2006, one would use the 2006 estimate (771,782) and the corresponding MOE at the 95-percent confidence level derived above (±14,723).

$$L_{95} = 771,782 - 14,723 = 757,059$$
$$U_{95} = 771,782 + 14,723 = 786,505$$

The 95-percent CI can thus be expressed as the range 757,059 to 786,505.

The CI is also useful when graphing estimates, to show the extent of sampling error present in the estimates, and for visually comparing estimates. For example, given the MOE at the 90-percent confidence level used in constructing the CI above, the user could be 90 percent certain that the value for the population was between 18,355 and 21,645. This CI can be represented visually as

$$\left(\underline{\hspace{6cm}}\right)$$
18,355   20,000   21,645

Coefficients of Variation

A coefficient of variation (CV) provides a measure of the relative amount of sampling error that is associated with a sample estimate. The CV is calculated as the ratio of the SE for an estimate to the estimate itself and is usually expressed as a percent. It is a useful barometer of the stability, and thus the usability of a sample estimate. It can also help a user decide whether a single-year or multiyear estimate should be used for analysis. The method for obtaining the SE for an estimate was described earlier.

The CV is a function of the overall sample size and the size of the population of interest. In general, as the estimation period increases, the sample size increases and therefore the size of the CV decreases. A small CV indicates that the sampling error is small relative to the estimate, and thus the user can be more confident that the estimate is close to the population value. In some applications a small CV for an estimate is desirable and use of a multiyear estimate will therefore be preferable to the use of a 1-year estimate that doesn't meet this desired level of precision.

For example, if an estimate of 20,000 had an SE of 1,000, then the CV for the estimate would be 5 percent ([1,000 /20,000] x 100). In terms of usability, the estimate is very reliable. If the CV was noticeably larger, the usability of the estimate could be greatly diminished.

While it is true that estimates with high CVs have important limitations, they can still be valuable as building blocks to develop estimates for higher levels of aggregation. Combining estimates across geographic areas or collapsing characteristic detail can improve the reliability of those estimates as evidenced by reductions in the CVs.

Calculating Coefficients of Variation From Standard Errors

The CV can be expressed as

$$CV = \frac{SE}{\hat{X}} \times 100$$

where $\hat{X}$ is the ACS estimate and SE is the derived SE for the ACS estimate.

For example, to determine the CV for the estimated number of civilian veterans in the state of Virginia in 2006, one would use the 2006 estimate (771,782), and the SE derived previously (7,512).

$$CV = \frac{7,512}{771,782} \times 100 = 0.1\%$$

This means that the amount of sampling error present in the estimate is only one-tenth of 1 percent the size of the estimate.

The text box below summarizes the formulas used when deriving alternative sampling error measures from the margin or error published with ACS estimates.

Deriving Sampling Error Measures From Published MOE
Margin Error (MOE) for Alternate Confidence Levels
$$MOE_{95} = \frac{1.960}{1.645} MOE_{ACS}$$ $$MOE_{99} = \frac{2.576}{1.645} MOE_{ACS}$$
Standard Error (SE)
$$SE = \frac{MOE_{ACS}}{1.645}$$
Confidence Interval (CI)
$$CI_{CL} = \left(\hat{X} - MOE_{CL}, \hat{X} + MOE_{CL} \right)$$
Coefficient of Variation (CV)
$$CV = \frac{SE}{\hat{X}} \times 100$$

Calculating Margins of Error for Derived Estimates

One of the benefits of being familiar with ACS data is the ability to develop unique estimates called derived estimates. These derived estimates are usually based on aggregating estimates across geographic areas or population subgroups for which combined estimates are not published in American FactFinder (AFF) tables (e.g., aggregate estimates for a three-county area or for four age groups not collapsed).

ACS tabulations provided through AFF contain the associated confidence intervals (pre-2005) or margins of error (MOEs) (2005 and later) at the 90-percent confidence level. However, when derived estimates are generated (e.g., aggregated estimates, proportions, or ratios not available in AFF), the user must calculate the MOE for these derived estimates. The MOE helps protect against misinterpreting small or nonexistent differences as meaningful.

MOEs calculated based on information provided in AFF for the components of the derived estimates will be at the 90-percent confidence level. If an MOE with a confidence level other than 90 percent is desired, the user should first calculate the MOE as instructed below and then convert the results to an MOE for the desired confidence level as described earlier in this appendix.

Calculating MOEs for Aggregated Count Data

To calculate the MOE for aggregated count data:
1) Obtain the MOE of each component estimate.
2) Square the MOE of each component estimate.
3) Sum the squared MOEs.
4) Take the square root of the sum of the squared MOEs.

The result is the MOE for the aggregated count. Algebraically, the MOE for the aggregated count is calculated as:

$$MOE_{agg} = \pm\sqrt{\sum_{c} MOE_c^2}$$

where MOE_c is the MOE of the cth component estimate.

The example below shows how to calculate the MOE for the estimated total number of females living alone in the three Virginia counties/independent cities that border Washington, DC (Fairfax and Arlington counties, Alexandria city) from the 2006 ACS.

Table 1. Data for Example 1

Characteristic	Estimate	MOE
Females living alone in Fairfax County (Component 1)	52,354	±3,303
Females living alone in Arlington County (Component 2)	19,464	±2,011
Females living alone in Alexandria city (Component 3)	17,190	±1,854

The aggregate estimate is:

$$\hat{X} = \hat{X}_{Fairfax} + \hat{X}_{Arlington} + \hat{X}_{Alexandria} =$$
$$52,354 + 19,464 + 17,190 = 89,008$$

Obtain MOEs of the component estimates:

$$MOE_{Fairfax} = \pm 3,303 ,$$
$$MOE_{Arlington} = \pm 2,011 ,$$
$$MOE_{Alexandria} = \pm 1,854$$

Calculate the MOE for the aggregate estimated as the square root of the sum of the squared MOEs.

$$MOE_{agg} = \pm\sqrt{(3,303)^2 + (2,011)^2 + (1,854)^2} =$$
$$\pm\sqrt{18,391,246} = \pm 4,289$$

Thus, the derived estimate of the number of females living alone in the three Virginia counties/independent cities that border Washington, DC, is 89,008, and the MOE for the estimate is ±4,289.

Calculating MOEs for Derived Proportions

The numerator of a proportion is a subset of the denominator (e.g., the proportion of single person households that are female). To calculate the MOE for derived proportions, do the following:

1) Obtain the MOE for the numerator and the MOE for the denominator of the proportion.
2) Square the derived proportion.
3) Square the MOE of the numerator.
4) Square the MOE of the denominator.
5) Multiply the squared MOE of the denominator by the squared proportion.
6) Subtract the result of (5) from the squared MOE of the numerator.
7) Take the square root of the result of (6).
8) Divide the result of (7) by the denominator of the proportion.

The result is the MOE for the derived proportion. Algebraically, the MOE for the derived proportion is calculated as:

$$MOE_p = \frac{\pm \sqrt{MOE_{num}^2 - (\hat{p}^2 * MOE_{den}^2)}}{\hat{X}_{den}}$$

where MOE_{num} is the MOE of the numerator.

MOE_{den} is the MOE of the denominator.

$\hat{p} = \dfrac{\hat{X}_{num}}{\hat{X}_{den}}$ is the derived proportion.

$\hat{X}_{num}$ is the estimate used as the numerator of the derived proportion.

$\hat{X}_{den}$ is the estimate used as the denominator of the derived proportion.

There are rare instances where this formula will fail—the value under the square root will be negative. If that happens, use the formula for derived ratios in the next section which will provide a conservative estimate of the MOE.

The example below shows how to derive the MOE for the estimated proportion of Black females 25 years of age and older in Fairfax County, Virginia, with a graduate degree based on the 2006 ACS.

Table 2. **Data for Example 2**		
Characteristic	Estimate	MOE
Black females 25 years and older with a graduate degree (numerator)	4,634	±989
Black females 25 years and older (denominator)	31,713	±601

The estimated proportion is:

$$\hat{p} = \frac{\hat{X}_{gradBF}}{\hat{X}_{BF}} = \frac{4,634}{31,713} = 0.1461$$

where $\hat{X}_{gradBF}$ is the ACS estimate of Black females 25 years of age and older in Fairfax County with a graduate degree and $\hat{X}_{BF}$ is the ACS estimate of Black females 25 years of age and older in Fairfax County.

Obtain MOEs of the numerator (number of Black females 25 years of age and older in Fairfax County with a graduate degree) and denominator (number of Black females 25 years of age and older in Fairfax County).

$$MOE_{num} = \pm 989, \quad MOE_{den} = \pm 601$$

Multiply the squared MOE of the denominator by the squared proportion and subtract the result from the squared MOE of the numerator.

$$MOE_{num}^2 - (\hat{p}^2 * MOE_{den}^2) =$$
$$(989)^2 - [(0.1461)^2 * (601)^2] =$$
$$978,121 - 7,712.3 = 970,408.7$$

Calculate the MOE by dividing the square root of the prior result by the denominator.

$$MOE_p = \frac{\pm \sqrt{970,408.7}}{31,373} = \frac{\pm 985.1}{31,373} = \pm 0.0311$$

Thus, the derived estimate of the proportion of Black females 25 years of age and older with a graduate degree in Fairfax County, Virginia, is 0.1461, and the MOE for the estimate is ±0.0311.

Calculating MOEs for Derived Ratios

The numerator of a ratio is not a subset (e.g., the ratio of females living alone to males living alone). To calculate the MOE for derived ratios:

1) Obtain the MOE for the numerator and the MOE for the denominator of the ratio.
2) Square the derived ratio.
3) Square the MOE of the numerator.
4) Square the MOE of the denominator.
5) Multiply the squared MOE of the denominator by the squared ratio.
6) Add the result of (5) to the squared MOE of the numerator.
7) Take the square root of the result of (6).
8) Divide the result of (7) by the denominator of the ratio.

The result is the MOE for the derived ratio. Algebraically, the MOE for the derived ratio is calculated as:

$$MOE_R = \frac{\pm \sqrt{MOE_{num}^2 + (\hat{R}^2 * MOE_{den}^2)}}{\hat{X}_{den}}$$

where MOE_{num} is the MOE of the numerator.

MOE_{den} is the MOE of the denominator.

$\hat{R} = \dfrac{\hat{X}_{num}}{\hat{X}_{den}}$ is the derived ratio.

$\hat{X}_{num}$ is the estimate used as the numerator of the derived ratio.

$\hat{X}_{den}$ is the estimate used as the denominator of the derived ratio.

The example below shows how to derive the MOE for the estimated ratio of Black females 25 years of age and older in Fairfax County, Virginia, with a graduate degree to Black males 25 years and older in Fairfax County with a graduate degree, based on the 2006 ACS.

Table 3. **Data for Example 3**		
Characteristic	Estimate	MOE
Black females 25 years and older with a graduate degree (numerator)	4,634	±989
Black males 25 years and older with a graduate degree (denominator)	6,440	±1,328

The estimated ratio is:

$$\hat{R} = \frac{\hat{X}_{gradBF}}{\hat{X}_{gradBM}} = \frac{4,634}{6,440} = 0.7200$$

Obtain MOEs of the numerator (number of Black females 25 years of age and older with a graduate degree in Fairfax County) and denominator (number of Black males 25 years of age and older in Fairfax County with a graduate degree).

$$MOE_{num} = \pm 989, \ MOE_{den} = \pm 1,328$$

Multiply the squared MOE of the denominator by the squared proportion and add the result to the squared MOE of the numerator.

$$MOE_{num}^2 + (\hat{R}^2 * MOE_{den}^2) =$$
$$(989)^2 + [(0.7200)^2 * (1,328)^2] =$$
$$978,121 + 913,318.1 = 1,891,259.1$$

Calculate the MOE by dividing the square root of the prior result by the denominator.

$$MOE_R = \frac{\pm \sqrt{1,891,259.1}}{6,440} = \frac{\pm 1,375.2}{6,440} = \pm 0.2135$$

Thus, the derived estimate of the ratio of the number of Black females 25 years of age and older in Fairfax County, Virginia, with a graduate degree to the number of Black males 25 years of age and older in Fairfax County, Virginia, with a graduate degree is 0.7200, and the MOE for the estimate is ±0.2135.

Calculating MOEs for the Product of Two Estimates

To calculate the MOE for the product of two estimates, do the following:

1) Obtain the MOEs for the two estimates being multiplied together.

2) Square the estimates and their MOEs.

3) Multiply the first squared estimate by the second estimate's squared MOE.

4) Multiply the second squared estimate by the first estimate's squared MOE.

5) Add the results from (3) and (4).

6) Take the square root of (5).

The result is the MOE for the product. Algebraically, the MOE for the product is calculated as:

$$MOE_{A \times B} = \pm \sqrt{A^2 \times MOE_B^2 + B^2 \times MOE_A^2}$$

where *A* and *B* are the first and second estimates, respectively.

MOE_A is the MOE of the first estimate.

MOE_B is the MOE of the second estimate.

The example below shows how to derive the MOE for the estimated number of Black workers 16 years and over in Fairfax County, Virginia, who used public transportation to commute to work, based on the 2006 ACS.

Table 4. **Data for Example 4**		
Characteristic	Estimate	MOE
Black workers 16 years and over (first estimate)	50,624	±2,423
Percent of Black workers 16 years and over who commute by public transportation (second estimate)	13.4%	±2.7%

To apply the method, the proportion (0.134) needs to be used instead of the percent (13.4). The estimated product is 50,624 × 0.134 = 6,784. The MOE is calculated by:

$$MOE_{A \times B} = \pm \sqrt{50,624^2 \times 0.027^2 + 0.134^2 \times 2,423^2}$$
$$= \pm 1,405$$

Thus, the derived estimate of Black workers 16 years and over who commute by public transportation is 6,784, and the MOE of the estimate is ±1,405.

Calculating MOEs for Estimates of "Percent Change" or "Percent Difference"

The "percent change" or "percent difference" between two estimates (for example, the same estimates in two different years) is commonly calculated as

$$Percent\ Change = 100\% * \frac{\hat{X}_2 - \hat{X}_1}{\hat{X}_1}$$

Because $\hat{X}_2$ is not a subset of $\hat{X}_1$, the procedure to calculate the MOE of a ratio discussed previously should be used here to obtain the MOE of the percent change.

The example below shows how to calculate the margin of error of the percent change using the 2006 and 2005 estimates of the number of persons in Maryland who lived in a different house in the U.S. 1 year ago.

Table 5. **Data for Example 5**		
Characteristic	Estimate	MOE
Persons who lived in a different house in the U.S. 1 year ago, 2006	802,210	±22,866
Persons who lived in a different house in the U.S. 1 year ago, 2005	762,475	±22,666

The percent change is:

$$Percent\ Change = 100\% * \frac{\hat{X}_2 - \hat{X}_1}{\hat{X}_1} =$$

$$100\% * \left(\frac{802,210 - 762,475}{762,475} \right) = 5.21\%$$

For use in the ratio formula, the ratio of the two estimates is:

$$\hat{R} = \frac{\hat{X}_2}{\hat{X}_1} = \frac{802,210}{762,475} = 1.0521$$

The MOEs for the numerator ($\hat{X}_2$) and denominator ($\hat{X}_1$) are:

$$MOE_2 = +/-22,866,\ MOE_1 = +/-22,666$$

Add the squared MOE of the numerator (MOE_2) to the product of the squared ratio and the squared MOE of the denominator (MOE_1):

$$MOE_2^2 + (\hat{R}^2 * MOE_1^2) =$$

$$(22,866)^2 + [(1.0521)^2 * (22,666)^2] =$$

$$1,091,528,529$$

Calculate the MOE by dividing the square root of the prior result by the denominator ($\hat{X}_1$).

$$MOE_R = \frac{\pm\sqrt{1,091,528,529}}{762,475} = \frac{\pm 33,038.3}{762,475} = \pm 0.0433$$

Finally, the MOE of the percent change is the MOE of the ratio, multiplied by 100 percent, or 4.33 percent.

The text box below summarizes the formulas used to calculate the margin of error for several derived estimates.

Calculating Margins of Error for Derived Estimates

Aggregated Count Data

$$MOE_{agg} = \pm\sqrt{\sum_c MOE_c^2}$$

Derived Proportions

$$MOE_p = \frac{\pm\sqrt{MOE_{num}^2 - (\hat{p}^2 * MOE_{den}^2)}}{\hat{X}_{den}}$$

Derived Ratios

$$MOE_R = \frac{\pm\sqrt{MOE_{num}^2 + (\hat{R}^2 * MOE_{den}^2)}}{\hat{X}_{den}}$$

U.S. Census Bureau, A Compass for Understanding and Using American Community Survey Data

Appendix 4.
Making Comparisons

One of the most important uses of the ACS estimates is to make comparisons between estimates. Several key types of comparisons are of general interest to users: 1) comparisons of estimates from different geographic areas within the same time period (e.g., comparing the proportion of people below the poverty level in two counties); 2) comparisons of estimates for the same geographic area across time periods (e.g., comparing the proportion of people below the poverty level in a county for 2006 and 2007); and 3) comparisons of ACS estimates with the corresponding estimates from past decennial census samples (e.g., comparing the proportion of people below the poverty level in a county for 2006 and 2000).

A number of conditions must be met when comparing survey estimates. Of primary importance is that the comparison takes into account the sampling error associated with each estimate, thus determining whether the observed differences between estimates are statistically significant. Statistical significance means that there is statistical evidence that a true difference exists within the full population, and that the observed difference is unlikely to have occurred by chance due to sampling. A method for determining statistical significance when making comparisons is presented in the next section. Considerations associated with the various types of comparisons that could be made are also discussed.

Determining Statistical Significance

When comparing two estimates, one should use the test for significance described below. This approach will allow the user to ascertain whether the observed difference is likely due to chance (and thus is not statistically significant) or likely represents a true difference that exists in the population as a whole (and thus is statistically significant).

The test for significance can be carried out by making several computations using the estimates and their corresponding standard errors (SEs). When working with ACS data, these computations are simple given the data provided in tables in the American FactFinder.

1) Determine the SE for each estimate (for ACS data, SE is defined by the positive value of the margin of error (MOE) divided by 1.645).[4]
2) Square the resulting SE for each estimate.
3) Sum the squared SEs.
4) Calculate the square root of the sum of the squared SEs.

---

[4] NOTE: If working with ACS single-year estimates for 2005 or earlier, use the value 1.65 rather than 1.645.

5) Calculate the difference between the two estimates.
6) Divide (5) by (4).
7) Compare the absolute value of the result of (6) with the critical value for the desired level of confidence (1.645 for 90 percent, 1.960 for 95 percent, 2.576 for 99 percent).
8) If the absolute value of the result of (6) is greater than the critical value, then the difference between the two estimates can be considered statistically significant at the level of confidence corresponding to the critical value used in (7).

Algebraically, the significance test can be expressed as follows:

If $\left| \dfrac{\hat{X}_1 - \hat{X}_2}{\sqrt{SE_1^2 + SE_2^2}} \right| > Z_{CL}$, then the difference

between estimates $\hat{X}_1$ and $\hat{X}_2$ is statistically significant at the specified confidence level, CL

where $\hat{X}_i$ is estimate i (=1,2)

SE_i is the SE for the estimate i (=1,2)

Z_{CL} is the critical value for the desired confidence level (=1.645 for 90 percent, 1.960 for 95 percent, 2.576 for 99 percent).

The example below shows how to determine if the difference in the estimated percentage of households in 2006 with one or more people of age 65 and older between State A (estimated percentage =22.0, SE=0.12) and State B (estimated percentage =21.5, SE=0.12) is statistically significant. Using the formula above:

$$\left| \frac{\hat{X}_1 - \hat{X}_2}{\sqrt{SE_1^2 + SE_2^2}} \right| = \left| \frac{22.0 - 21.5}{\sqrt{(0.12)^2 + (0.12)^2}} \right| =$$

$$\left| \frac{0.5}{\sqrt{0.015 + 0.015}} \right| = \left| \frac{0.5}{\sqrt{0.03}} \right| = \left| \frac{0.5}{0.173} \right| = 2.90$$

Since the test value (2.90) is greater than the critical value for a confidence level of 99 percent (2.576), the difference in the percentages is statistically significant at a 99-percent confidence level. This is also referred to as statistically significant at the alpha = 0.01 level. A rough interpretation of the result is that the user can be 99 percent certain that a difference exists between the percentages of households with one or more people aged 65 and older between State A and State B.

By contrast, if the corresponding estimates for State C and State D were 22.1 and 22.5, respectively, with standard errors of 0.20 and 0.25, respectively, the formula would yield

$$\left| \frac{\hat{X}_1 - \hat{X}_2}{\sqrt{SE_1^2 + SE_2^2}} \right| = \left| \frac{22.5 - 22.1}{\sqrt{(0.20)^2 + (0.25)^2}} \right| =$$

$$\left| \frac{0.4}{\sqrt{0.04 + 0.0625}} \right| = \left| \frac{0.4}{\sqrt{0.1025}} \right| = \left| \frac{0.4}{0.320} \right| = 1.25$$

Since the test value (1.25) is less than the critical value for a confidence level of 90 percent (1.645), the difference in percentages is not statistically significant. A rough interpretation of the result is that the user cannot be certain to any sufficient degree that the observed difference in the estimates was not due to chance.

Comparisons Within the Same Time Period

Comparisons involving two estimates from the same time period (e.g., from the same year or the same 3-year period) are straightforward and can be carried out as described in the previous section. There is, however, one statistical aspect related to the test for statistical significance that users should be aware of. When comparing estimates within the same time period, the areas or groups will generally be nonoverlapping (e.g., comparing estimates for two different counties). In this case, the two estimates are independent, and the formula for testing differences is statistically correct.

In some cases, the comparison may involve a large area or group and a subset of the area or group (e.g., comparing an estimate for a state with the corresponding estimate for a county within the state or comparing an estimate for all females with the corresponding estimate for Black females). In these cases, the two estimates are not independent. The estimate for the large area is partially dependent on the estimate for the subset and, strictly speaking, the formula for testing differences should account for this partial dependence. However, unless the user has reason to believe that the two estimates are strongly correlated, it is acceptable to ignore the partial dependence and use the formula for testing differences as provided in the previous section. However, if the two estimates are positively correlated, a finding of statistical significance will still be correct, but a finding of a lack of statistical significance based on the formula may be incorrect. If it is important to obtain a more exact test of significance, the user should consult with a statistician about approaches for accounting for the correlation in performing the statistical test of significance.

Comparisons Across Time Periods

Comparisons of estimates from different time periods may involve different single-year periods or different multiyear periods of the same length within the same area. Comparisons across time periods should be made only with comparable time period estimates. Users are advised against comparing single-year estimates with multiyear estimates (e.g., comparing 2006 with 2007–2009) and against comparing multiyear estimates of differing lengths (e.g., comparing 2006–2008 with 2009–2014), as they are measuring the characteristics of the population in two different ways, so differences between such estimates are difficult to interpret. When carrying out any of these types of comparisons, users should take several other issues into consideration.

When comparing estimates from two different single-year periods, one prior to 2006 and the other 2006 or later (e.g., comparing estimates from 2005 and 2007), the user should recognize that from 2006 on the ACS sample includes the population living in group quarters (GQ) as well as the population living in housing units. Many types of GQ populations have demographic, social, or economic characteristics that are very different from the household population. As a result, comparisons between 2005 and 2006 and later ACS estimates could be affected. This is particularly true for areas with a substantial GQ population. For most population characteristics, the Census Bureau suggests users make comparisons across these time periods only if the geographic area of interest does not include a substantial GQ population. For housing characteristics or characteristics published only for the household population, this is obviously not an issue.

Comparisons Based on Overlapping Periods

When comparing estimates from two multiyear periods, ideally comparisons should be based on nonoverlapping periods (e.g., comparing estimates from 2006–2008 with estimates from 2009–2011). The comparison of two estimates for different, but overlapping periods is challenging since the difference is driven by the nonoverlapping years. For example, when comparing the 2005–2007 ACS with the 2006–2008 ACS, data for 2006 and 2007 are included in both estimates. Their contribution is subtracted out when the estimate of differences is calculated. While the interpretation of this difference is difficult, these comparisons can be made with caution. Under most circumstances, the estimate of difference should not be interpreted as a reflection of change between the last 2 years.

The use of MOEs for assessing the reliability of change over time is complicated when change is being evaluated using multiyear estimates. From a technical standpoint, change over time is best evaluated with multiyear estimates that do not overlap. At the same time,

Appendix **A-19**

many areas whose only source of data will be 5-year estimates will not want to wait until 2015 to evaluate change (i.e., comparing 2005–2009 with 2010–2014).

When comparing two 3-year estimates or two 5-year estimates of the same geography that overlap in sample years one must account for this sample overlap. Thus to calculate the standard error of this difference use the following approximation to the standard error:

$$SE(\hat{X}_1 - \hat{X}_2) \cong \sqrt{(1-C)}\sqrt{SE_1^{\,2} + SE_2^{\,2}}$$

where C is the fraction of overlapping years. For example, the periods 2005–2009 and 2007–2011 overlap for 3 out of 5 years, so C=3/5=0.6. If the periods do not overlap, such as 2005–2007 and 2008–2010, then C=0.

With this SE one can test for the statistical significance of the difference between the two estimates using the method outlined in the previous section with one modification; substitute $\sqrt{(1-C)}\sqrt{SE_1^{\,2} + SE_2^{\,2}}$ for $\sqrt{SE_1^{\,2} + SE_2^{\,2}}$ in the denominator of the formula for the significance test.

Comparisons With Census 2000 Data

In Appendix 2, major differences between ACS data and decennial census sample data are discussed. Factors such as differences in residence rules, universes, and reference periods, while not discussed in detail in this appendix, should be considered when comparing ACS estimates with decennial census estimates. For example, given the reference period differences, seasonality may affect comparisons between decennial census and ACS estimates when looking at data for areas such as college towns and resort areas.

The Census Bureau subject matter specialists have reviewed the factors that could affect differences between ACS and decennial census estimates and they have determined that ACS estimates are similar to those obtained from past decennial census sample data for most areas and characteristics. The user should consider whether a particular analysis involves an area or characteristic that might be affected by these differences.[5]

When comparing ACS and decennial census sample estimates, the user must remember that the decennial census sample estimates have sampling error associated with them and that the standard errors for both ACS and census estimates must be incorporated when performing tests of statistical significance. Appendix 3 provides the calculations necessary for determining

statistical significance of a difference between two estimates. To derive the SEs of census sample estimates, use the method described in Chapter 8 of either the Census 2000 Summary File 3 Technical Documentation <http://www.census.gov/prod/cen2000/doc/sf3.pdf> or the Census 2000 Summary File 4 Technical Documentation <http://www.census.gov/prod/cen2000/doc/sf4.pdf>.

A conservative approach to testing for statistical significance when comparing ACS and Census 2000 estimates that avoids deriving the SE for the Census 2000 estimate would be to assume the SE for the Census 2000 estimate is the same as that determined for the ACS estimate. The result of this approach would be that a finding of statistical significance can be assumed to be accurate (as the SE for the Census 2000 estimate would be expected to be less than that for the ACS estimate), but a finding of no statistical significance could be incorrect. In this case the user should calculate the census long-form standard error and follow the steps to conduct the statistical test.

Comparisons With 2010 Census Data

Looking ahead to the 2010 decennial census, data users need to remember that the socioeconomic data previously collected on the long form during the census will not be available for comparison with ACS estimates. The only common variables for the ACS and 2010 Census are sex, age, race, ethnicity, household relationship, housing tenure, and vacancy status.

The critical factor that must be considered when comparing ACS estimates encompassing 2010 with the 2010 Census is the potential impact of housing and population controls used for the ACS. As the housing and population controls used for 2010 ACS data will be based on the Population Estimates Program where the estimates are benchmarked on the Census 2000 counts, they will not agree with the 2010 Census population counts for that year. The 2010 population estimates may differ from the 2010 Census counts for two major reasons—the true change from 2000 to 2010 is not accurately captured by the estimates and the completeness of coverage in the 2010 Census is different than coverage of Census 2000. The impact of this difference will likely affect most areas and states, and be most notable for smaller geographic areas where the potential for large differences between the population controls and the 2010 Census population counts is greater.

Comparisons With Other Surveys

Comparisons of ACS estimates with estimates from other national surveys, such as the Current Population Survey, may be of interest to some users. A major consideration in making such comparisons will be that ACS

---

[5] Further information concerning areas and characteristics that do not fit the general pattern of comparability can be found on the ACS Web site at <http://www.census.gov/acs/www/UseData/compACS.htm>.

estimates include data for populations in both institutional and noninstitutional group quarters, and estimates from most national surveys do not include institutional populations. Another potential for large effects when comparing data from the ACS with data from other national surveys is the use of different questions for measuring the same or similar information.

Sampling error and its impact on the estimates from the other survey should be considered if comparisons and statements of statistical difference are to be made, as described in Appendix 3. The standard errors on estimates from other surveys should be derived according to technical documentation provided for those individual surveys.

Finally, the user wishing to compare ACS estimates with estimates from other national surveys should consider the potential impact of other factors, such as target population, sample design and size, survey period, reference period, residence rules, and interview modes on estimates from the two sources.

Appendix 5.

Using Dollar-Denominated Data

Dollar-denominated data refer to any characteristics for which inflation adjustments are used when producing annual estimates. For example, income, rent, home value, and energy costs are all dollar-denominated data.

Inflation will affect the comparability of dollar-denominated data across time periods. When ACS multiyear estimates for dollar-denominated data are generated, amounts are adjusted using inflation factors based on the Consumer Price Index (CPI).

Given the potential impact of inflation on observed differences of dollar-denominated data across time periods, users should adjust for the effects of inflation. Such an adjustment will provide comparable estimates accounting for inflation. In making adjustments, the Census Bureau recommends using factors based on the All Items CPI-U-RS (CPI research series). The Bureau of Labor Statistics CPI indexes through 2006 are found at <http://www.bls.gov/cpi/cpiurs1978_2006.pdf>. Explanations follow.

Creating Single-Year Income Values

ACS income values are reported based on the amount of income received during the 12 months preceding the interview month. This is the income reference period. Since there are 12 different income reference periods throughout an interview year, 12 different income inflation adjustments are made. Monthly CPI-U-RSs are used to inflation-adjust the 12 reference period incomes to a single reference period of January through December of the interview year. Note that there are no inflation adjustments for single-year estimates of rent, home value, or energy cost values.

Adjusting Single-Year Estimates Over Time

When comparing single-year income, rent, home value, and energy cost value estimates from two different years, adjustment should be made as follows:

1) Obtain the All Items CPI-U-RS Annual Averages for the 2 years being compared.

2) Calculate the inflation adjustment factor as the ratio of the CPI-U-RS from the more recent year to the CPI-U-RS from the earlier year.

3) Multiply the dollar-denominated data estimated for the earlier year by the inflation adjustment factor.

The inflation-adjusted estimate for the earlier year can be expressed as:

$$\hat{X}_{Y1,Adj} = \frac{CPI_{Y2}}{CPI_{Y1}} \hat{X}_{Y1}$$

where CPI_{Y1} is the All Items CPI-U-RS Annual Average for the earlier year (Y1).

CPI_{Y2} is the All Items CPI-U-RS Annual Average for the more recent year (Y2).

$\hat{X}_{Y1}$ is the published ACS estimate for the earlier year (Y1).

The example below compares the national median value for owner-occupied mobile homes in 2005 ($37,700) and 2006 ($41,000). First adjust the 2005 median value using the 2005 All Items CPI-U-RS Annual Average (286.7) and the 2006 All Items CPI-U-RS Annual Average (296.1) as follows:

$$\hat{X}_{2005,Adj} = \frac{296.1}{286.7} \times \$37,700 = \$38,936$$

Thus, the comparison of the national median value for owner-occupied mobile homes in 2005 and 2006, in 2006 dollars, would be $38,936 (2005 inflation-adjusted to 2006 dollars) versus $41,000 (2006 dollars).

Creating Values Used in Multiyear Estimates

Multiyear income, rent, home value, and energy cost values are created with inflation adjustments. The Census Bureau uses the All Items CPI-U-RS Annual Averages for each year in the multiyear time period to calculate a set of inflation adjustment factors. Adjustment factors for a time period are calculated as ratios of the CPI-U-RS Annual Average from its most recent year to the CPI-U-RS Annual Averages from each of its earlier years. The ACS values for each of the earlier years in the multiyear period are multiplied by the appropriate inflation adjustment factors to produce the inflation-adjusted values. These values are then used to create the multiyear estimates.

As an illustration, consider the time period 2004–2006, which consisted of individual reference-year income values of $30,000 for 2006, $20,000 for 2005, and $10,000 for 2004. The multiyear income components are created from inflation-adjusted reference period income values using factors based on the All Items CPI-U-RS Annual Averages of 277.4 (for 2004), 286.7 (for 2005), and 296.1 (for 2006). The adjusted 2005 value is the ratio of 296.1 to 286.7 applied to $20,000, which equals $20,656. Similarly, the 2004 value is the ratio of 296.1 to 277.4 applied to $10,000, which equals $10,674.

Adjusting Multiyear Estimates Over Time

When comparing multiyear estimates from two different time periods, adjustments should be made as follows:

1) Obtain the All Items CPI-U-RS Annual Average for the most current year in each of the time periods being compared.

2) Calculate the inflation adjustment factor as the ratio of the CPI-U-RS Annual Average in (1) from the most recent year to the CPI-U-RS in (1) from the earlier years.

3) Multiply the dollar-denominated estimate for the earlier time period by the inflation adjustment factor.

The inflation-adjusted estimate for the earlier years can be expressed as:

$$\hat{X}_{P1,Adj} = \frac{CPI_{P2}}{CPI_{P1}} \hat{X}_{P1}$$

where CPI_{P1} is the All Items CPI-U-RS Annual Average for the last year in the earlier time period (P1).

CPI_{P2} is the All Items CPI-U-RS Annual Average for the last year in the most recent time period (P2).

$\hat{X}_{P1}$ is the published ACS estimate for the earlier time period (P1).

As an illustration, consider ACS multiyear estimates for the two time periods of 2001–2003 and 2004–2006. To compare the national median value for owner-occupied mobile homes in 2001–2003 ($32,000) and 2004–2006 ($39,000), first adjust the 2001–2003 median value using the 2003 All Items CPI-U-RS Annual Averages (270.1) and the 2006 All Items CPI-U-RS Annual Averages (296.1) as follows:

$$\hat{X}_{2001-2003,Adj} = \frac{296.1}{270.1} \times \$32,000 = \$35,080$$

Thus, the comparison of the national median value for owner-occupied mobile homes in 2001–2003 and 2004–2006, in 2006 dollars, would be $35,080 (2001–2003 inflation-adjusted to 2006 dollars) versus $39,000 (2004–2006, already in 2006 dollars).

Issues Associated With Inflation Adjustment

The recommended inflation adjustment uses a national level CPI and thus will not reflect inflation differences that may exist across geographies. In addition, since the inflation adjustment uses the All Items CPI, it will not reflect differences that may exist across characteristics such as energy and housing costs.

Appendix 6.

Measures of Nonsampling Error

All survey estimates are subject to both sampling and nonsampling error. In Appendix 3, the topic of sampling error and the various measures available for understanding the uncertainty in the estimates due to their being derived from a sample, rather than from an entire population, are discussed. The margins of error published with ACS estimates measure only the effect of sampling error. Other errors that affect the overall accuracy of the survey estimates may occur in the course of collecting and processing the ACS, and are referred to collectively as nonsampling errors.

Broadly speaking, nonsampling error refers to any error affecting a survey estimate outside of sampling error. Nonsampling error can occur in complete censuses as well as in sample surveys, and is commonly recognized as including coverage error, unit nonresponse, item nonresponse, response error, and processing error.

Types of Nonsampling Errors

Coverage error occurs when a housing unit or person does not have a chance of selection in the sample (undercoverage), or when a housing unit or person has more than one chance of selection in the sample, or is included in the sample when they should not have been (overcoverage). For example, if the frame used for the ACS did not allow the selection of newly constructed housing units, the estimates would suffer from errors due to housing undercoverage.

The final ACS estimates are adjusted for under- and overcoverage by controlling county-level estimates to independent total housing unit controls and to independent population controls by sex, age, race, and Hispanic origin (more information is provided on the coverage error definition page of the "ACS Quality Measures" Web site at <http://www.census.gov/acs/www /UseData/sse/cov/cov_def.htm>). However, it is important to measure the extent of coverage adjustment by comparing the precontrolled ACS estimates to the final controlled estimates. If the extent of coverage adjustments is large, there is a greater chance that differences in characteristics of undercovered or overcovered housing units or individuals differ from those eligible to be selected. When this occurs, the ACS may not provide an accurate picture of the population prior to the coverage adjustment, and the population controls may not eliminate or minimize that coverage error.

Unit nonresponse is the failure to obtain the minimum required information from a housing unit or a resident of a group quarter in order for it to be considered a completed interview. Unit nonresponse means that no survey data are available for a particular sampled unit

or person. For example, if no one in a sampled housing unit is available to be interviewed during the time frame for data collection, unit nonresponse will result.

It is important to measure unit nonresponse because it has a direct effect on the quality of the data. If the unit nonresponse rate is high, it increases the chance that the final survey estimates may contain bias, even though the ACS estimation methodology includes a nonresponse adjustment intended to control potential unit nonresponse bias. This will happen if the characteristics of nonresponding units differ from the characteristics of responding units.

Item nonresponse occurs when a respondent fails to provide an answer to a required question or when the answer given is inconsistent with other information. With item nonresponse, while some responses to the survey questionnaire for the unit are provided, responses to other questions are not obtained. For example, a respondent may be unwilling to respond to a question about income, resulting in item nonresponse for that question. Another reason for item nonresponse may be a lack of understanding of a particular question by a respondent.

Information on item nonresponse allows users to judge the completeness of the data on which the survey estimates are based. Final estimates can be adversely impacted when item nonresponse is high, because bias can be introduced if the actual characteristics of the people who do not respond to a question differ from those of people who do respond to it. The ACS estimation methodology includes imputations for item nonresponse, intended to reduce the potential for item nonresponse bias.

Response error occurs when data are reported or recorded incorrectly. Response errors may be due to the respondent, the interviewer, the questionnaire, or the survey process itself. For example, if an interviewer conducting a telephone interview incorrectly records a respondent's answer, response error results. In the same way, if the respondent fails to provide a correct response to a question, response error results. Another potential source of response error is a survey process that allows proxy responses to be obtained, wherein a knowledgeable person within the household provides responses for another person within the household who is unavailable for the interview. Even more error prone is allowing neighbors to respond.

Processing error can occur during the preparation of the final data files. For example, errors may occur if data entry of questionnaire information is incomplete

or inaccurate. Coding of responses incorrectly also results in processing error. Critical reviews of edits and tabulations by subject matter experts are conducted to keep errors of this kind to a minimum.

Nonsampling error can result in random errors and systematic errors. Of greatest concern are systematic errors. Random errors are less critical since they tend to cancel out at higher geographic levels in large samples such as the ACS.

On the other hand, systematic errors tend to accumulate over the entire sample. For example, if there is an error in the questionnaire design that negatively affects the accurate capture of respondents' answers, processing errors are created. Systematic errors often lead to a bias in the final results. Unlike sampling error and random error resulting from nonsampling error, bias caused by systematic errors cannot be reduced by increasing the sample size.

ACS Quality Measures

Nonsampling error is extremely difficult, if not impossible, to measure directly. However, the Census Bureau has developed a number of indirect measures of nonsampling error to help inform users of the quality of the ACS estimates: sample size, coverage rates, unit response rates and nonresponse rates by reason, and item allocation rates. Starting with the 2007 ACS, these measures are available in the B98 series of detailed tables on AFF. Quality measures for previous years are available on the "ACS Quality Measures" Web site at <http://www.census.gov/acs/www/UseData/sse/>.

Sample size measures for the ACS summarize information for the housing unit and GQ samples. The measures available at the state level are:[6]

    Housing units
        Number of initial addresses selected
        Number of final survey interviews

    Group quarters people (beginning with the 2006 ACS)
        Number of initial persons selected
        Number of final survey interviews

Sample size measures may be useful in special circumstances when determining whether to use single-year or multiyear estimates in conjunction with estimates of the population of interest. While the coefficient of variation (CV) should typically be used to determine usability, as explained in Appendix 3, there may be some situations where the CV is small but the user has reason to believe the sample size for a subgroup is very small and the robustness of the estimate is in question.

For example, the Asian-alone population makes up roughly 1 percent (8,418/656,700) of the population in Jefferson County, Alabama. Given that the number of successful housing unit interviews in Jefferson County for the 2006 ACS were 4,072 and assuming roughly 2.5 persons per household (or roughly 12,500 completed person interviews), one could estimate that the 2006 ACS data for Asians in Jefferson County are based on roughly 150 completed person interviews.

Coverage rates are available for housing units, and total population by sex at both the state and national level. Coverage rates for total population by six race/ethnicity categories and the GQ population are also available at the national level. These coverage rates are a measure of the extent of adjustment to the survey weights required during the component of the estimation methodology that adjusts to population controls. Low coverage rates are an indication of greater potential for coverage error in the estimates.

Unit response and nonresponse rates for housing units are available at the county, state, and national level by reason for nonresponse: refusal, unable to locate, no one home, temporarily absent, language problem, other, and data insufficient to be considered an interview. Rates are also provided separately for persons in group quarters at the national and state levels.

A low unit response rate is an indication that there is potential for bias in the survey estimates. For example, the 2006 housing unit response rates are at least 94 percent for all states. The response rate for the District of Columbia in 2006 was 91 percent.

Item allocation rates are determined by the content edits performed on the individual raw responses and closely correspond to item nonresponse rates. Overall housing unit and person characteristic allocation rates are available at the state and national levels, which combine many different characteristics. Allocation rates for individual items may be calculated from the B99 series of imputation detailed tables available in AFF.

Item allocation rates do vary by state, so users are advised to examine the allocation rates for characteristics of interest before drawing conclusions from the published estimates.

---

[6] The sample size measures for housing units (number of initial addresses selected and number of final survey interviews) and for group quarters people cannot be used to calculate response rates. For the housing unit sample, the number of initial addresses selected includes addresses that were determined not to identify housing units, as well as initial addresses that are subsequently subsampled out in preparation for personal visit nonresponse follow-up. Similarly, the initial sample of people in group quarters represents the expected sample size within selected group quarters prior to visiting and sampling of residents.

Appendix 7.

Implications of Population Controls on ACS Estimates

As with most household surveys, the American Community Survey data are controlled so that the numbers of housing units and people in categories defined by age, sex, race, and Hispanic origin agree with the Census Bureau's official estimates. The American Community Survey (ACS) measures the characteristics of the population, but the official count of the population comes from the previous census, updated by the Population Estimates Program.

In the case of the ACS, the total housing unit estimates and the total population estimates by age, sex, race and Hispanic origin are controlled at the county (or groups of counties) level. The group quarters total population is controlled at the state level by major type of group quarters. Such adjustments are important to correct the survey data for nonsampling and sampling errors. An important source of nonsampling error is the potential under-representation of hard-to-enumerate demographic groups. The use of the population controls results in ACS estimates that more closely reflect the level of coverage achieved for those groups in the preceding census. The use of the population estimates as controls partially corrects demographically implausible results from the ACS due to the ACS data being based on a sample of the population rather than a full count. For example, the use of the population controls "smooths out" demographic irregularities in the age structure of the population that result from random sampling variability in the ACS.

When the controls are applied to a group of counties rather than a single county, the ACS estimates and the official population estimates for the individual counties may not agree. There also may not be agreement between the ACS estimates and the population estimates for levels of geography such as subcounty areas where the population controls are not applied.

The use of population and housing unit controls also reduces random variability in the estimates from year to year. Without the controls, the sampling variability in the ACS could cause the population estimates to increase in one year and decrease in the next (especially for smaller areas or demographic groups), when the underlying trend is more stable. This reduction in variability on a time series basis is important since results from the ACS may be used to monitor trends over time. As more current data become available, the time series of estimates from the Population Estimates Program are revised back to the preceding census while the ACS estimates in previous years are not. Therefore, some differences in the ACS estimates across time may be due to changes in the population estimates.

For single-year ACS estimates, the population and total housing unit estimates for July 1 of the survey year are used as controls. For multiyear ACS estimates, the controls are the average of the individual year population estimates.

Appendix 8.
Other ACS Resources

Background and Overview Information

American Community Survey Web Page Site Map: <http://www.census.gov/acs/www/Site_Map.html> This link is the site map for the ACS Web page. It provides an overview of the links and materials that are available online, including numerous reference documents.

What Is the ACS? <http://www.census.gov/acs/www/SBasics/What/What1.htm> This Web page includes basic information about the ACS and has links to additional information including background materials.

ACS Design, Methodology, Operations

American Community Survey Design and Methodology Technical Paper: <http://www.census.gov/acs/www/Downloads/tp67.pdf> This document describes the basic design of the 2005 ACS and details the full set of methods and procedures that were used in 2005. Please watch our Web site as a revised version will be released in the fall of 2008, detailing methods and procedures used in 2006 and 2007.

About the Data (Methodology: <http://www.census.gov/acs/www/AdvMeth/> This Web page contains links to information on ACS data collection and processing, evaluation reports, multiyear estimates study, and related topics.

ACS Quality

Accuracy of the Data (2007): <http://www.census.gov/acs/www/Downloads/ACS/accuracy2007.pdf> This document provides data users with a basic understanding of the sample design, estimation methodology, and accuracy of the 2007 ACS data.

ACS Sample Size: <http://www.census.gov/acs/www/SBasics/SSizes/SSizes06.htm> This link provides sample size information for the counties that were published in the 2006 ACS. The initial sample size and the final completed interviews are provided. The sample sizes for all published counties and county equivalents starting with the 2007 ACS will only be available in the B98 series of detailed tables on American FactFinder.

ACS Quality Measures: <http://www.census.gov/acs/www/UseData/sse/> This Web page includes information about the steps taken by the Census Bureau to improve the accuracy of ACS data. Four indicators of survey quality are described and measures are provided at the national and state level.

Guidance on Data Products and Using the Data

How to Use the Data: <http://www.census.gov/acs/www/UseData/> This Web page includes links to many documents and materials that explain the ACS data products.

Comparing ACS Data to other sources: <http://www.census.gov/acs/www/UseData/compACS.htm> Tables are provided with guidance on comparing the 2007 ACS data products to 2006 ACS data and Census 2000 data.

Fact Sheet on Using Different Sources of Data for Income and Poverty: <http://www.census.gov/hhes/www/income/factsheet.html> This fact sheet highlights the sources that should be used for data on income and poverty, focusing on comparing the ACS and the Current Population Survey (CPS).

Public Use Microdata Sample (PUMS): <http://www.census.gov/acs/www/Products/PUMS/> This Web page provides guidance in accessing ACS microdata.