A Compass for Understanding and Using American Community Survey Data

What Federal Agencies Need to Know

Issued
December 2008

U S C E N S U S B U R E A U

Helping You Make Informed Decisions

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU

United States Census 2010

Acknowledgments

Frederick J. Eggers, Former Deputy Assistant Secretary for Economic Affairs, U.S. Department of Housing and Urban Development, drafted this handbook for the U.S. Census Bureau's American Community Survey Office. **Kennon R. Copeland** and **John H. Thompson** of National Opinion Research Center at the University of Chicago drafted the technical appendixes. **Edward J. Spar**, Executive Director, Council of Professional Associations on Federal Statistics, **Frederick J. Cavanaugh**, Executive Business Director, Sabre Systems, Inc., **Susan P. Love**, Consultant, **Linda A. Jacobsen**, Vice President, Domestic Programs, Population Reference Bureau, and **Mark Mather**, Associate Vice President, Domestic Programs, Population Reference Bureau, provided initial review of this handbook.

Deborah H. Griffin, Special Assistant to the Chief of the American Community Survey Office, provided the concept and directed the development and release of a series of handbooks entitled *A Compass for Understanding and Using American Community Survey Data.* **Cheryl V. Chambers**, **Colleen D. Flannery**, **Cynthia Davis Hollingsworth**, **Susan L. Hostetter**, **Pamela M. Klein**, **Anna M. Owens**, **Clive R. Richmond**, **Enid Santana**, and **Nancy K. Torrieri** contributed to the planning and review of this handbook series.

The American Community Survey program is under the direction of **Arnold A. Jackson**, Associate Director for Decennial Census, **Daniel H. Weinberg**, Assistant Director for the American Community Survey and Decennial Census, and **Susan Schechter**, Chief, American Community Survey Office.

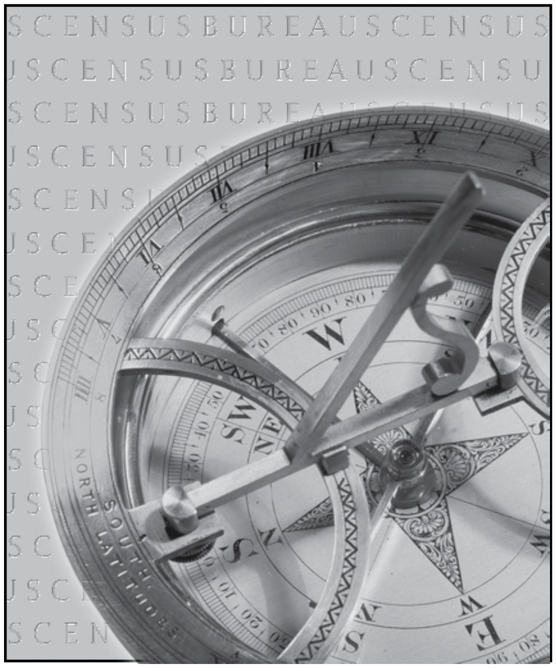
Other individuals who contributed to the review and release of these handbooks include **Dee Alexander**, **Herman Alvarado**, **Mark Asiala**, **Frank Ambrose**, **Maryam Asi**, **Arthur Bakis**, **Genora Barber**, **Michael Beaghen**, **Judy Belton**, **Lisa Blumerman**, **Scott Boggess**, **Ellen Jean Bradley**, **Stephen Buckner**, **Whittona Burrell**, **Edward Castro**, **Gary Chappell**, **Michael Cook**, **Russ Davis**, **Carrie Dennis**, **Jason Devine**, **Joanne Dickinson**, **Barbara Downs**, **Maurice Eleby**, **Sirius Fuller**, **Dale Garrett**, **Yvonne Gist**, **Marjorie Hanson**, **Greg Harper**, **William Hazard**, **Steve Hefter**, **Douglas Hillmer**, **Frank Hobbs**, **Todd Hughes**, **Trina Jenkins**, **Nicholas Jones**, **Anika Juhn**, **Donald Keathley**, **Wayne Kei**, **Karen King**, **Debra Klein**, **Vince Kountz**, **Ashley Landreth**, **Steve Laue**, **Van Lawrence**, **Michelle Lowe**, **Maria Malagon**, **Hector Maldonado**, **Ken Meyer**, **Louisa Miller**, **Stanley Moore**, **Alfredo Navarro**, **Timothy Olson**, **Dorothy Paugh**, **Marie Pees, Marc Perry**, **Greg Pewett**, **Roberto Ramirez**, **Dameka Reese**, **Katherine Reeves**, **Lil Paul Reyes**, **Patrick Rottas**, **Merarys Rios**, **J. Gregory Robinson**, **Anne Ross**, **Marilyn Sanders**, **Nicole Scanniello**, **David Sheppard**, **Joanna Stancil**, **Michael Starsinic**, **Lynette Swopes**, **Anthony Tersine**, **Carrie Werner**, **Edward Welniak**, **Andre Williams**, **Steven Wilson**, **Kai Wu**, and **Matthew Zimolzak**.

Linda Chen and **Amanda Perry** of the Administrative and Customer Services Division, **Francis Grailand Hall**, Chief, provided publications management, graphics design and composition, and editorial review for the print and electronic media. **Claudette E. Bennett**, Assistant Division Chief, and **Wanda Cevis**, Chief, Publications Services Branch, provided general direction and production management.

A Compass for Understanding and Using American Community Survey Data

What Federal Agencies Need to Know

Issued December 2008



U.S. Department of Commerce
Carlos M. Gutierrez,
Secretary

John J. Sullivan,
Deputy Secretary

Economics and Statistics Administration
Cynthia A. Glassman,
Under Secretary for Economic Affairs

U.S. CENSUS BUREAU
Steve H. Murdock,
Director

Suggested Citation

U.S. Census Bureau,
A Compass for Understanding and Using American Community Survey Data: What Federal Agencies Need to Know
U.S. Government Printing Office, Washington, DC, 2008.

Economics and Statistics Administration

Cynthia A. Glassman,
Under Secretary for Economic Affairs

U.S. CENSUS BUREAU

Steve H. Murdock,
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

Foreword

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

High school teachers

Business community

Researchers

Federal agencies

Media

Congress

Puerto Rico Community Survey data users (in Spanish)

Public Use Microdata Sample (PUMS) data users

Users of data for rural areas

State and local governments

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.

1. What Is the American Community Survey (ACS)?

While the first decennial census took place in 1790, only since 1940 has the U.S. Census Bureau collected data for the decennial census in two ways: (1) a set of questions administered to *all* housing units in order to count the population and gather basic demographic information, such as age, sex, race, and ethnicity and (2) an *additional* set of questions administered to a *sample* of households and housing units in order to obtain more detailed demographic, housing, social, and economic information. In recent censuses, the set of questions asked of all people has become known as the "short form," while the additional questions asked only of a sample have become known as the "long form." What we know about households—their incomes, their education, their employment, their housing—at the state, county, city, and census tract levels—has usually come from the long-form sample.

In 2005, the Census Bureau launched the American Community Survey (ACS) to replace the functions performed by the long form. Prior to 2005, the Census Bureau conducted extensive testing of ACS methods in specific site tests and in a series of national tests. The ACS collects socioeconomic and housing information continuously from a national *sample* of housing units and people living in group quarters, and the Census Bureau tabulates these data on a calendar year basis. In August and September 2006, the Census Bureau released the 2005 ACS data for all states, Puerto Rico, the District of Columbia, and for specified geographic areas, including counties and cities, with total populations of at least 65,000. Similar data were released in August and September 2007 for the 2006 ACS. In 2010, the Census Bureau will begin releasing ACS data annually for areas ranging from states to census tracts and block groups. The ACS is an integral component of the decennial census program, eliminating the need for a long-form sample in the decennial census. There will *not* be a long form in the 2010 Census nor in any future decennial census.

Federal agencies have used decennial census data, particularly the detailed information collected on a sample basis, for a variety of purposes. Several examples include:

- **Eligibility determinations.** Agencies use census sample data to determine the eligibility of places for participation in federal programs. For example, the Low Income Housing Tax Credit provides extra tax credits to selected housing projects in certain qualified census tracts. These tracts have been identified based on a special tabulation of household income distributions at the census tract level.

- **Allocation of funds.** Federal agencies use long-form data to allocate funds. For example, the Department of Agriculture's Smith Lever Act and the Department of Education's Preschool Grants for Children with Disabilities program allocate funds to communities using funding formulas that rely on decennial census data.

- **Program parameters.** Federal programs often operate with parameters that vary by location. For example, the Department of Housing and Urban Development (HUD) uses decennial income data in its calculations of the income limits that determine at the local level the eligibility of households for various forms of housing assistance.

- **Justification for discretionary grants.** In reviewing applications for discretionary grants, agencies require applicants to justify the need for a grant, and applicants have often employed decennial census sample data in the justification. For instance, the Department of Transportation works with states and local governments to obtain a special tabulation of journey-to-work data at the census tract level for use by applicants in models that assess the impact of a proposed grant.

- **Policy analysis and evaluation.** Agencies rely on decennial sample data to guide policy development and facilitate research and evaluation.

In the past, federal agencies used the data from the long-form sample for these functions; from now on, agencies will rely on ACS data to carry out these functions.

2. Purpose of This Handbook

This handbook is intended to aid analysts, program administrators, and policy makers within federal agencies when they use the ACS in carrying out the business of their agencies. Because agencies depend on the detailed housing and socioeconomic data that have been provided by the decennial census long-form sample, the advent of the ACS is of great interest and possibly some concern to the analysts and policy makers in agencies that use these data. Data users will want to know:

- Whether the ACS will provide the same information as the long form.
- Whether ACS data will be as reliable as long-form data.
- How having annual data will affect their programs.
- How program recipients might view the shift from the long-form census sample to the ACS.

This handbook is designed to help users in federal agencies understand the ACS and answer these questions. While this handbook will describe the essential features of the ACS in detail and will discuss some important statistical and practical considerations, the bottom line is that federal agencies should be able to use ACS data for the same purposes for which they used decennial census sample data. Clearly, the more frequent updating of information available in the ACS changes the statistical framework in which federal agencies have been operating and opens new opportunities for agencies to advance their missions and base policy and decision-making on more current information.

The remainder of this handbook is organized as follows:

Section 3 compares the information the ACS furnishes with the information that agencies obtained in the past from the long-form sample of the decennial census.

Sections 4, 5, and 6 explain the essential characteristics of the ACS, provide some perspective on their importance, and suggest ways to deal with problems that might arise from specific features.

- Section 4 discusses how to interpret ACS estimates since they are based on data collected over periods of 1 year, 3 years, and 5 years.
- Section 5 introduces some technical issues that users should be aware of.
- Section 6 explores some concerns that may arise as federal agencies make the transition from using the decennial long-form sample data to using the ACS data.

The final two sections furnish information on how to obtain ACS information and what to expect from the ACS in the future.

- Section 7 explains how federal users can access ACS data in ways that facilitate their use in federal programs.
- Section 8 offers some insight into the future of the ACS.

Throughout this handbook, text boxes will contain additional information related to the discussion in the main text. A glossary at the back of the handbook provides definitions for key terms, and a set of appendixes offers more in-depth explorations of some of the issues discussed in the text.

The goal of this handbook is to give the reader the basics needed to access and use ACS data. The technical appendixes furnish more detailed information on important subjects. In addition, the Census Bureau has produced a number of reports on selected topics involving the interpretation and quality of ACS data. The handbook cites the technical appendixes and Census Bureau documents at appropriate places in the text, and Appendix 8 lists the most useful references.

3. How Does the ACS Compare With the Decennial Census Long Form?

Most agencies will use ACS data in applications that have—until recently—used data from the decennial census long-form sample. To facilitate the transition between these two data sources, this section provides an overview comparison of the ACS and the decennial sample. In content, design, geographical coverage, and products, the ACS closely resembles the decennial sample. The most important differences involve sample size, the time between data releases, and the relationship between geography and the timing of data releases.[1]

Content. The content of the ACS is similar to the content of the Census 2000 long form. In both cases, the Census Bureau worked with federal agencies and the Office of Management and Budget to ensure that agency legislative requirements were provided for and that the public was not burdened with any question that was not based on a legislative requirement or clear programmatic need. Table 1 summarizes the range of information collected by the ACS.

The questionnaires for the ACS and the Census 2000 long form include many of the same questions using very similar language. However, there are some differences. For example, the Census 2000 long form asked whether a resident lived in the same unit 5 years ago while the ACS asks whether a resident lived in the same unit 1 year ago. Also the Census 2000 long form asked about income in 1999 (the previous calendar year), while the ACS asks about income over the 12 months preceding the interview date.

QUESTIONNAIRES

To understand how similar the two surveys are, users should go online and compare the latest ACS questionnaire with the questionnaire used for the Census 2000 long form. The link to the questionnaire being used in the 2008 ACS is <http://www.census.gov/acs/www/AdvMeth/2008Quest.html> and the link to the Census 2000 long-form questionnaire is <http://www.census.gov/dmd/www/2000quest.html>.

Table 1. **Subjects Included in the American Community Survey**

Demographic Characteristics	Social Characteristics	Housing Characteristics
Age	Marital Status and Marital History*	Year Structure Built
Sex	Fertility	Units in Structure
Hispanic Origin	Grandparents as Caregivers	Year Moved Into Unit
Race	Ancestry	Rooms
Relationship to Householder (e.g., spouse)	Place of Birth, Citizenship, and Year of Entry	Bedrooms
	Language Spoken at Home	Kitchen Facilities
Economic Characteristics	Educational Attainment and School Enrollment	Plumbing Facilities
Income	Residence One Year Ago	House Heating Fuel
Food Stamps Benefit	Veteran Status, Period of Military Service, and VA Service-Connected Disability Rating*	Telephone Service Available
Labor Force Status	Disability	Farm Residence
Industry, Occupation, and Class of Worker		
Place of Work and Journey to Work		**Financial Characteristics**
Work Status Last Year		Tenure (Owner/Renter)
Vehicles Available		Housing Value
Health Insurance Coverage*		Rent
		Selected Monthly Owner Costs

*Marital History, VA Service-Connected Disability Rating, and Health Insurance Coverage are new for 2008.
Source: Subjects Planned for the 2010 Census and American Community Survey, U.S. Census Bureau.

[1] Appendix 2 provides a more detailed comparison of ACS data and decennial census sample data. Users can find a comprehensive description of the methods and procedures currently used in the ACS in *American Community Survey Design and Methodology Technical Paper* at <http://www.census.gov/acs/www/Downloads/tp67.pdf>.

Survey design. The Census Bureau maintains a nationwide list of addresses that it uses for mailing questionnaires for the decennial census, the ACS, and other surveys. For Census 2000, every housing unit on this list received either a short form or a long form. Since the long form incorporated all the questions from the short form, this meant that all people living in every housing unit were asked the questions on the short form. Only a sample of housing units received the Census 2000 long form. Residents of group quarters were also surveyed, using both the short form and the long form. Nearly all the Census 2000 forms were mailed or delivered in March 2000. The Census Bureau kept track of responses and used enumerators to follow-up on all housing units that did not complete and mail back their census questionnaire.

In contrast, using the nationwide address list, the Census Bureau mails ACS questionnaires to a sample of approximately 250,000 addresses every month and tries to obtain completed interviews from these addresses within 3 months. In the first month, the Census Bureau mails a survey questionnaire to the sampled addresses followed by a second questionnaire if the first is not returned. In the second month, if the Census Bureau has received no response to these questionnaires, it attempts to collect the information by telephone. Finally, in the third month, for those addresses from which there is still no response, the Census Bureau sends a highly trained interviewer to a subsample ranging from one-third to one-half of the nonresponding addresses in an area. The sampling rate for the ACS in-person interviews is higher in census tracts with a lower predicted percentage of mail and telephone respondents.

Geographical coverage. Starting in 2010, the Census Bureau will publish ACS data for the same geographies covered in the published long-form data from Census 2000. These geographies include, but are not limited to, the nation, states, counties, places, American Indian reservations, census tracts, and block groups. A listing of the major geographies for which the Census Bureau will publish ACS data is furnished below in the discussion of the relationship between geography and the timing of data release. (See Table 3.) Definitions of these geographies are provided in the glossary at the end of the handbook.

Time between data releases. A fundamental distinction between ACS data and data from the decennial long form is that new ACS data are released every year, not just once every 10 years. This basic difference has three important implications for federal agencies.

- Agencies will not have to wait 10 years between censuses to update key information. They will be working with the most current data.
- Using ACS data may require more resources. Analysis of long-form data within an agency normally took place once a decade. If agencies choose to take advantage of the more frequent availability of new ACS data, the agencies will need to analyze new data more often, perhaps annually.
- Shortening the time between releases of new data will call attention to the reliability of published estimates. The ACS and its predecessor, the decennial census long form, are sample surveys; therefore, estimates based on either source are subject to sampling error. (See explanation of sampling error later in this section.) However, users are more likely to blame sampling errors for observed differences if the time between observations is shorter. For example, a 2 percentage point change in the poverty rate may not seem unusual over a 10-year period, but the same 2 percentage point change in a single year may appear questionable even though sampling errors may cause both changes. For this reason, agencies can expect their clients to question changes in estimates more frequently than when the estimates were 10 years apart.

Geography and data release. Another important distinction between the ACS and the decennial long-form sample is how the Census Bureau reports on geographies of different sizes. As befits a once-a-decade survey, the Census Bureau released long-form data collected in 2000 for all levels of geography down to the block group. The ACS is a continuous survey, and the Census Bureau reports data when it believes that the sample for a particular location is sufficient to make the published estimates reasonably reliable. Table 2 summarizes how the Census Bureau will release ACS data in three different formats: 1-year estimates, 3-year estimates, and 5-year estimates.

The 1-year estimates began in 2006 with the release of ACS data collected in 2005. These data were released for geographic areas with populations of 65,000 and greater. The release of 3-year estimates began in 2008 based on ACS data collected from 2005 through 2007. Three-year estimates are produced for areas with populations of 20,000 and greater. The release of 5-year estimates begins in 2010 derived from ACS data collected from 2005 through 2009. These 5-year estimates are published for all geographic entities down to census tracts and block groups.

The 3-year and the 5-year estimates are called "period estimates" because of the way that the Census Bureau constructs them. Each year the Census Bureau creates new multiyear estimates by adding the newest year of ACS data and dropping the oldest year of ACS data. For example, the 5-year estimates released in 2010 will be based on data collected in 2005, 2006, 2007,

Table 2. Collection and Release Dates of Data From the ACS (Pattern Repeats After 2010)

Type of period estimate	Release year (late summer or fall)						
	2006	2007	2008	2009	2010	2011	2012
	Calendar year(s) of monthly data collection						
1-year period estimate for areas with 65,000 or more people	2005	2006	2007	2008	2009	2010	2011
3-year period estimates for areas with 20,000 or more people			2005–2007	2006–2008	2007–2009	2008–2010	2009–2011
5-year period estimates for all areas including those with fewer than 20,000 people					2005–2009	2006–2010	2007–2011

Source: Citro and Kalton (2007: Table 2-6, page 50).

2008, and 2009. These estimates will be called the 2005–2009 ACS 5-year estimates. The 5-year estimates released in 2011 will be based on data collected in 2006, 2007, 2008, 2009, and 2010. These estimates will be called the 2006–2010 ACS 5-year estimates. The difference between the 2005–2009 estimates and the 2006–2010 estimates is that the latter will include 2010 data and exclude 2005 data. Both estimates will include the same data from 2006, 2007, 2008, and 2009.

Table 3 summarizes the types of geographic areas receiving only 5-year estimates, 5-year and 3-year estimates, and areas receiving all three data sets.

Table 3 has several important implications for users of ACS data:

- As the number of areas in column 1 makes clear, the vast majority of areas will receive only the 5-year estimates.

- Estimates based on a single year of ACS data will be available only for geographic areas with the largest populations. Nearly all states, congressional districts, and Public Use Microdata Areas receive 1-year estimates.

- Smaller geographic areas, including census tracts and census block groups, will have data that portray conditions in those locations not in a specific year but over several years. Census tract and block group data are only produced as 5-year estimates.

- For some levels of geography, the Census Bureau will release up to three sets of estimates every year. For example, in 2010, the ACS Web site will include three sets of the same tables for Birmingham, Alabama, (a city with 208,000 residents)—one set using data from 2009, one set using data from the 2007–2009 period, and one set using data from the 2005–2009 period.

In Section 4, this handbook will examine each of these implications more fully.

Products. Section 7 will explain how users can obtain ACS data online. The tables available from the American FactFinder (AFF) are similar, but not identical, to what was used for Census 2000 long-form data. The Census Bureau has refined these tables over time in response to data user needs. In general, ACS tables include the same topics as did Census 2000 and these tables have similar patterns of rows and columns. The tables cover the subjects listed in Table 1.

Sample Size. The ACS (and the long form in Census 2000 and prior years) obtains its information from a sample of housing units and group quarters. The Census 2000 long form was distributed to approximately 1 in 6 housing units. The ACS uses an annual sample of approximately 1 in 40 housing units but combines information from successive years so that eventually its published tables will represent approximately 1 in 40, 3 in 40, or 5 in 40 (1 in 8) housing units. In the published tables, the Census Bureau transforms the ACS sample information into estimates of the number of people, households, and housing units. While the tables include numbers that some users interpret as "counts," it is critical to remember that these numbers are estimates, not counts.

Sampling error. Questions regarding the precision of estimates apply to both census long form and ACS data. However, the small sample sizes in the ACS as

Table 3. ACS Publication Plans by Size of Geographic Area

Type of geographic area	Total number of areas	Percent of total areas receiving . . . 1-year, 3-year, & 5-year estimates	3-year & 5-year estimates only	5-year estimates only
States and District of Columbia	51	100.0	0.0	0.0
Congressional districts	435	100.0	0.0	0.0
Public Use Microdata Areas*	2,071	99.9	0.1	0.0
Metropolitan statistical areas	363	99.4	0.6	0.0
Micropolitan statistical areas	576	24.3	71.2	4.5
Counties and county equivalents	3,141	25.0	32.8	42.2
Urban areas	3,607	10.4	12.9	76.7
School districts (elementary, secondary, and unified)	14,120	6.6	17.0	76.4
American Indian areas, Alaska Native areas, and Hawaiian homelands	607	2.5	3.5	94.1
Places (cities, towns, and census designated places)	25,081	2.0	6.2	91.8
Townships and villages (minor civil divisions)	21,171	0.9	3.8	95.3
ZIP Code tabulation areas	32,154	0.0	0.0	100.0
Census tracts	65,442	0.0	0.0	100.0
Census block groups	208,801	0.0	0.0	100.0

* When originally designed, each PUMA contained a population of about 100,000. Over time, some of these PUMAs have gained or lost population. However, due to the population displacement in the greater New Orleans areas caused by Hurricane Katrina in 2005, Louisiana PUMAs 1801, 1802, and 1805 no longer meet the 65,000-population threshold for 1-year estimates. With reference to Public Use Microdata Sample (PUMS) data, records for these PUMAs were combined to ensure ACS PUMS data for Louisiana remain complete and additive.

Source: U.S. Census Bureau, 2008. This tabulation is restricted to geographic areas in the United States. It was based on the population sizes of geographic areas from the July 1, 2007, Census Bureau Population Estimates and geographic boundaries as of January 1, 2007. Because of the potential for changes in population size and geographic boundaries, the actual number of areas receiving 1-year, 3-year, and 5-year estimates may differ from the numbers in this table.

compared with the census long-form sample result in more statistical variation in ACS estimates than in corresponding long-form estimates. The Census Bureau provides users with information (the “margin of error”) on the statistical accuracy of every estimate in ACS tables.[2] For example, the 2006 ACS reported that there were 20,153 people 25 years and older in Birmingham, Alabama, who have a bachelor’s degree and that the margin of sampling error associated with this estimate is plus or minus 2,588. Since the margin of sampling error provided represents the 90-percent confidence level, this means that there is a probability of 90 percent or more that the number of people with a bachelor’s degree in Birmingham, Alabama, in 2006 is between 17,565 (20,153 – 2,588) and 22,741 (20,153 + 2,588).

[2] Appendix 3 explains various measures of statistical reliability and how to interpret them.

Relating data to time and place. There are some subtle but important differences that users should note when comparing ACS data with similar data from the decennial long form. The Census Bureau collected Census 2000 long-form data in a narrow timeframe, mostly in March through June of 2000, whereas it collects ACS data continuously throughout every year. Timing differences affect the interpretation of variables, even when the questions are worded identically. For example, both Census 2000 and the ACS ask about whether each household member was employed “last week.” The Census 2000 answers were used to calculate an unemployment rate for a period around April 1, 2000, while the ACS answers are used to calculate an unemployment rate representative of the entire time period over which the ACS data are or will be collected. The continuous nature of data collection in the ACS also leads the Census Bureau to adjust for changes that take place over time. Prior to publication, the Census Bureau adjusts some data reported in dollars to account for changes in the real value of money. It also adjusts the geography for its data products to agree with the most recent changes in the boundaries of the area being described. Section 4 discusses these important features of ACS data.

4. How to Interpret and Use ACS Data

This section and Section 5 discuss a number of issues that federal agencies should be aware of when using ACS data. In most cases, the issues are generic, that is, they will have similar implications for different agencies.

Time and ACS Data

The preceding section pointed out several distinctions between ACS and census long-form data, among them that the ACS's continuous data collection produces estimates with a different interpretation than estimates from once-a-decade decennial data. In general, ACS data describe conditions over the time period during which the Census Bureau collected the data (1, 3, or 5 years); Census 2000 data described conditions around April 1, 2000. Thus, ACS data provide "time-period" estimates, while prior long-form data provided approximate "point-in-time" estimates. The time-period nature of ACS data has several important consequences.

Interpretation of 1-year estimates. ACS estimates represent the "average" of conditions over the time period during which the Census Bureau collected the data. Saying that ACS estimates are "averages" seems simple and straightforward, but an "average" may mean different things depending on the variable being studied. For example, educational attainment is a population characteristic that changes slowly. In this case, the ACS estimate is likely to provide a close approximation to conditions at any one point during the year even if the Census Bureau collected the data over the course of the year. Unemployment, on the other hand, fluctuates throughout the year. The ACS unemployment estimate therefore may not be a close approximation to the unemployment rate on any given day during the year. Users must remember that, for 1-year ACS data, the estimates do not represent conditions at any one point during a year but do reflect an average over the year.[3]

Interpretation of multiyear estimates. The multiyear ACS estimates have the same interpretation as single-year estimates, but, in these cases, the aggregated estimates represent information collected over 3 or 5 years rather than 1 year.[4] The longer data collection period means that there is more time for conditions to change during the course of measurement. In interpreting multiyear estimates, users should take into account the variable, the geographical area, and the length of the data collection period. In particular, users should not assume that a multiyear estimate represents conditions at the end of the multiyear period. Two metropolitan areas may have the same 5-year average poverty rates but the poverty rates of the two metropolitan areas may be substantially different on the last day of the measurement period.

Comparison of multiyear estimates. Federal agencies and other users will benefit from having new information on conditions in counties, cities, and census tracts every year. Users will want to compare data from different years to see how conditions may have changed. Comparisons involving single-year ACS estimates are straightforward because they involve comparisons of data collected in two independent samples. Comparisons involving multiyear ACS data are complicated by the fact that the estimates being compared may be based on some of the same data.

It is important to distinguish between comparisons involving overlapping and nonoverlapping samples. The top panel of Table 4 illustrates how the six consecutive 3-year estimates that will be released between 2008 and 2013 relate to one another. The 2005–2007 3-year estimates *overlap* with the 2006–2008 and 2007–2009 3-year period estimates because all three estimates use data collected in 2007. The 2005–2007 3-year estimates and the 2008–2010 3-year estimates are said to be *nonoverlapping* because they use none of the same data. The bottom panel of Table 4 compares the six consecutive 5-year estimates that will be released between 2010 and 2015. The 5-year estimates that will be released in 2010 will overlap with the 5-year estimates that will be released in 2011 through 2014. In Table 4, only the 2010 and 2015 releases will be nonoverlapping. This table shows that starting with the data collected in 2009, every ACS interview is used to produce nine different sets of estimates: one 1-year estimate, three 3-year estimates, and five 5-year estimates.

Comparisons involving overlapping estimates depict changes only as calendar year data sets are dropped and added to the sample. Any changes that may have occurred between the beginning and the end of the survey period are muted by the use of the same

[3] Appendix 1 discusses in more detail how to interpret 1-year ACS estimates.

[4] Appendix 1 discusses in more detail how to interpret multiyear ACS estimates.

information (the overlapping samples) in both estimates. For example, changes between the 2005–2007 3-year estimates and the 2006–2008 3-year estimates are entirely driven by the changes between 2005 and 2008, because the 2006 and 2007 data are used in both estimates.

Overlapping estimates may give users a sense of underlying trends but, because the number of observations dropped and added between samples may be small, a suspected trend may be the result of the deletion and addition of a relatively small number of atypical cases. Nonoverlapping estimates will give users the best indication of whether and how conditions may have changed. Thus, rather than comparing estimates from the 2005–2007 3-year estimate with the 2006–2008 3-year estimate, agencies should compare the 2005–2007 3-year estimate with the 2008–2010 3-year estimate. Data users will need to be patient until nonoverlapping data sets are available. Typical statistical tests cannot be applied to overlapping samples because the overlap gives a false sense of the size of the samples being compared.[5]

[5] Refer to Appendix 4 for guidance on comparing ACS estimates with overlapping time periods.

Table 4. **Overlapping and Nonoverlapping Multiyear Estimates**

	Consecutive 3-year Period Estimates and the Years in Which the ACS Data Used in Their Construction Were Collected					
	2005–2007 3-year estimates	**2006–2008 3-year estimates**	**2007–2009 3-year estimates**	**2008–2010 3-year estimates**	**2009–2011 3-year estimates**	**2010–2012 3-year estimates**
2005	collected					
2006	collected ➜	reused				
2007	collected ➜	reused ➜	reused			
2008	*released*	collected ➜	reused ➜	reused		
2009		*released*	collected ➜	reused ➜	reused	
2010			*released*	collected ➜	reused ➜	reused
2011				*released*	collected ➜	reused
2012					*released*	collected
2013						*released*
	Consecutive 5-year Period Estimates and the Years in Which the ACS Data Used in Their Construction Were Collected					
	2005–2009 5-year estimates	**2006–2010 5-year estimates**	**2007–2011 5-year estimates**	**2008–2012 5-year estimates**	**2009–2013 5-year estimates**	**2010–2014 5-year estimates**
2005	collected					
2006	collected ➜	reused				
2007	collected ➜	reused ➜	reused			
2008	collected ➜	reused ➜	reused ➜	reused		
2009	collected ➜	reused ➜	reused ➜	reused ➜	reused	
2010	*released*	collected ➜	reused ➜	reused ➜	reused ➜	reused
2011		*released*	collected ➜	reused ➜	reused ➜	reused
2012			*released*	collected ➜	reused ➜	reused
2013				*released*	collected ➜	reused
2014					*released*	collected
2015						*released*

Source: U.S. Census Bureau.

Implications. There are four important implications of this discussion.

- First, users of ACS data should understand that an average is not a precise description of conditions at any one point in time, just as users of Census 2000 long-form data realized that a point-in-time estimate was not a precise description of conditions over a time period.

- Second, users should keep the distinction between time-period and point-in-time estimates in mind when comparing ACS data with 2000 or earlier census long-form data. While the tables from the two sources may use the same or similar wording, the numbers reported refer to different concepts. For example, the 2005 1-year ACS unemployment rate refers to unemployment during 2005 for an area while the 2000 decennial census unemployment rate refers to unemployment around April 1, 2000, for an area.

- Third, once users understand that ACS data represent conditions over specific periods of time, they should have no concerns about using ACS data to compare conditions across places or across time periods because the data are collected and reported consistently for all areas and for all survey years.[6]

- Fourth, over-time comparisons that rely on nonoverlapping samples will provide the clearest indications of how conditions may have changed.

More Than One Estimate for the Same Measurement

The Census Bureau releases ACS estimates in more than one format, depending upon the size of the area being reported. This feature presents users in federal agencies with choices not available from the decennial long form. This subsection discusses these choices.

Choosing Between 1-Year, 3-Year, and 5-Year Estimates

By 2010, analysts and policy makers in federal agencies will have three alternative ACS data sets to choose from for states, counties, and places with 65,000 or more residents. If, for example, in 2010, an agency wants to know the median household size for Salinas, California, a city with 150,724 inhabitants in 2000, it can use the 1-year estimate based on 2009 data, the 3-year estimate based on data from 2007 through 2009, or the 5-year estimate based on data from 2005 through 2009. As discussed next, which estimate to use depends upon several factors.

Intended use. Comparisons with other places should be based on the information derived from the same sets of estimates.[7] If an agency wants to compare Salinas with other large cities, any of the three sets of estimates would be appropriate, since all three estimates would be available for every large city. But, if the agency wanted to compare Salinas with smaller cities, then the 3-year or 5-year estimates should be used depending upon the size of the smallest city in the comparison. The Census Bureau has recommended that users should not compare 1-year data with 3- or 5-year data. Small cities with populations under 20,000 would have only 5-year estimates, so the agency should choose 5-year estimates for all the cities being compared. If the agency wanted to compare Salinas in 2010 to Salinas in 2011, then the nonoverlapping samples provided by the 1-year data would provide the clearest comparison. For over-time comparisons involving areas that do not have 1-year estimates, agencies should use nonoverlapping estimates. This means that the smaller cities would have only two or three nonoverlapping data points to compare during a decade.

Desired precision. The 5-year estimates are based on larger samples than the 3-year estimates, and the 3-year estimates have larger samples than the 1-year estimates. Larger samples lead to smaller margins of error and produce more precise estimates. The margins of error for the single-year samples are approximately 1.73 times larger than the margins of error for the 3-year estimates and approximately 2.24 times larger than the margins of error for the 5-year estimates. Some population and housing characteristics reported by the ACS are found infrequently in the population or housing stock; for these characteristics, use of the multiyear estimates is strongly recommended because of the relatively small number of cases with these characteristics in the 1-year estimates. The margins of error reported in the ACS give agencies the information needed to assess the precision of the alternative estimates. Refer to Appendix 1 for additional information on precision and the use of single-year versus multiyear estimates.

Currency of the estimates. If an agency wants information on current conditions, then 1-year estimates, if available, would be preferred to 3-year estimates and 3-year estimates would be preferred to 5-year estimates.

Implication. How an agency chooses between the available ACS estimates will depend upon its perception of how to balance currency, precision, and consistency to meet the needs of its programs.

[6] This presumes that the survey questionnaire remains constant over time. Section 8 will discuss how the Census Bureau plans to incorporate modifications to the questionnaire to minimize disruptions in intertemporal comparisons.

[7] Appendix 4 discusses comparisons between ACS estimates.

Choosing Between ACS Estimates and Alternative Federal Estimates

For some important measures of socioeconomic and housing conditions, the ACS is not the only source of data produced by the federal government. Previously, estimates derived from decennial long-form samples never competed with information from these other sources because the alternative sources could furnish updated statistics throughout the decade between censuses. The ACS, on the other hand, can deliver updated estimates annually and, therefore, may be a plausible alternative to these other series. The choice between using the ACS and alternative federal statistical series hinges on several factors that agencies must take into account in the context of their program needs.

Conceptual accuracy. The ACS, like the previous decennial long form, uses a questionnaire that is designed to collect a wide range of demographic, economic, and housing information. This breadth of coverage limits the depth to which the ACS questionnaire can probe any one issue. Other federal surveys contain more questions related directly to their primary focus, which in most cases allows them to offer fuller descriptions of the phenomenon they are describing and to incorporate important conceptual refinements into their measurements. The Current Population Survey (CPS) is an excellent example. Because the CPS's main mission is to measure conditions in the labor market, it contains numerous questions to ascertain whether unemployed individuals are actively seeking employment. Conceptually, only people who are in the labor force—that is, either employed or actively seeking employment—should be included when calculating the unemployment rate. Because the long form did not have all the questions that the CPS has on labor force participation, it historically produced higher estimates of the unemployment rate than the estimates from the CPS covering the same time period. The ACS is subject to the same limitations on depth of coverage as the decennial census long form.

Sample size. The ACS initially selects approximately 3 million housing unit addresses each year, resulting in about 2 million interviews—a sample size that dwarfs the sample size of any other federal household survey and enables the Census Bureau to publish substate estimates.[8]

Geographic detail. Sample size differences become increasingly important at finer levels of geography. Few federal surveys create estimates for counties and places with populations as low as 65,000, the counties and places for which the ACS furnishes 1-year estimates. The precision of the ACS 1-year estimates varies by the size of the area and can be easily determined

[8] Due to the use of subsampling in the final stage of data collection, the final annual sample of completed interviews is about 2 million housing units.

OTHER IMPORTANT FEDERAL DATA SERIES

The Bureau of Labor Statistics (BLS) produces monthly estimates of employment, unemployment, and labor force participation at the national, state, and city levels. BLS uses the Current Population Survey for its national and state unemployment estimates; it uses statistical models to provide estimates of labor force, employment, unemployment, and the unemployment rate for metropolitan areas and cities down to 10,000 population. Information on the national data can be found at <http://www.bls.gov/cps/>. The estimates of local employment conditions can be found at <http://www.bls.gov/lau/>.

The Current Population Survey/Housing Vacancy Survey produces quarterly estimates of homeownership and vacancy rates at the national and regional levels. Information on these housing series and the most recent and historical data can be found at <http://www.census.gov/hhes/www/housing/hvs/hvs.html>.

The Department of Housing and Urban Development's (HUD's) American Housing Survey publishes information on the condition of the housing stock at the national and regional level every 2 years and for selected metropolitan areas on a 6-year cycle. The report on the 2007 national American Housing Survey is available at <http://www.census.gov/prod/2008pubs/h150-07.pdf>.

The Census Bureau's annual publications contain detailed information on income and poverty at the national level. The 2007 report is available at <http://www.census.gov/hhes/www/poverty/poverty07.html>.

from the accuracy information released by the Census Bureau with the ACS estimates.

Timeliness and frequency of reporting. As noted earlier, the ACS 1-year estimates measure the condition over the preceding year, and the Census Bureau releases these typically in August or September of the following year. Other federal surveys provide data on unemployment on a monthly basis, usually within a month of data collection. They also provide data on homeownership and housing vacancies on a quarterly basis, usually within 2 months of the end of the quarter. The housing data from the American Housing Survey are reported only every other year and normally 6 months after the year in which they were collected. The Census Bureau times the release of the ACS data on income and poverty to coincide with the release of its report on this topic.

Implication. Agencies will have to balance various concerns in deciding whether to use ACS estimates or an alternative federal statistical series. The relevant concerns will be conceptual accuracy, geographical detail, statistical precision at lower levels of geography, and the timeliness and frequency of reporting.

5. Technical Issues Involving ACS Data

This section deals with several more technical features of ACS data that users in federal agencies should know about.

Household Population and the Group Quarters (GQ) Population

Like the Census 2000 long-form sample, the ACS distinguishes between people living in housing units (households) and people living in group quarters (GQ). Group quarters include university housing, nursing homes, psychiatric hospitals, military barracks or ships, religious housing, and correctional facilities.[9] Homeless people who sleep in shelters are included in the GQ population surveyed by the ACS, but the "street" homeless are not. Neither the Census 2000 long-form sample nor the ACS furnishes information on the street homeless. The ACS began collecting data from samples of the GQ population in 2006; only the household population was surveyed for the 2005 ACS. The ACS typically includes both the household and GQ populations in tables that describe the characteristics of the total population but includes only the household population in tables that describe the characteristics of households or housing units. Currently, only national-level ACS estimates are released by major GQ type. PUMS files, discussed in Section 7, allow users to tabulate the characteristics of the GQ population separately from those of the household population at the PUMA level. While such tabulations are possible, users need

HOW NOT TO INTERPRET INFLATION-ADJUSTED VARIABLES

Over the time periods for which ACS data are tabulated (1-year estimates, 3-year estimates, and 5-year estimates), variables denominated in dollars change because the individual variables are subject to economic trends. Incomes generally increase over time, as do rents and housing values. These increases reflect both general inflation and other trends such as increases in productivity or the rising cost of land. The Census Bureau's inflation adjustment does not adjust for underlying trends in these variables. Instead, the adjustment is designed to put the collected data into dollars of equal purchasing power.

For example, consider a household surveyed by the ACS in April 2005 and reporting a rent of $610. The inflation adjustment for the 3-year ACS estimate covering 2005–2007 would calculate how many dollars would be needed in 2007 to purchase the equivalent of general goods and services that cost $610 in 2005—in this case, roughly $648 would be needed. In this example, $610 is adjusted for general inflation; the adjustment does not consider either rental market trends in rents that might arise from increases in the size and quality of rental units or other factors.

[9] For a complete listing of types of group quarters, see 2006 ACS group quarters type codes and definitions at <http://www.census.gov/acs/www/UseData/GQ/def.htm>.

to consider sampling error and other factors when they produce GQ tabulations from the PUMS files.

Dollar-Valued Data and Adjustments for Inflation

A number of ACS variables are reported as dollar values; these variables include income and its various components, rent, and property value. For its multiyear estimates, the Census Bureau will adjust many, but not all, of the variables that are measured in dollar terms. Income and its components are the only variables adjusted for inflation in both 1-year and multiyear estimates. Users of ACS data should consult the "Subject Definitions" document available on the ACS Web site for the appropriate year of the ACS survey to determine whether the ACS made an adjustment for inflation.[10] Whenever the Census Bureau makes an adjustment for inflation, it uses the Consumer Price Index Research Series (CPI-U-RS) published by the Bureau of Labor Statistics to transform the estimates into *real dollars as of the reference year—which is the year that interviews were conducted or the most current year for multiyear estimates*. Refer to Appendix 5 for more information about inflation adjustments.

The ACS sometimes combines or otherwise manipulates the information provided by respondents to create useful measures; these measures are called *composite* or *derived* variables. For example, the ACS reports the estimated number of individuals and families living in poverty; it also provides estimates of renter households by the ratio of gross rent to income. The classification of individuals, families, and households for these tables are made based on the data received *without* adjustment for inflation. For example, if a completed questionnaire is received in July 2007, the Census Bureau will calculate the ratio of gross rent to income by adding the components of gross rent as they were reported on the questionnaire and dividing by the income reported on the questionnaire. This procedure defines the composite variable at the time of data collection, an approach consistent with the correct interpretation of the composite variables.

Geographic Boundary Changes

The Census Bureau organizes and reports ACS data by geography; it releases tables for the nation, states, counties, places, and other entities with physical boundaries. Since the boundaries of places and other entities can and do change over the period in which the Census Bureau collects ACS data, a consistent rule is needed on how to report data when boundaries change.

The Census Bureau always uses the legal boundary in effect on January 1 of the year in which it tabulates the data. Table 5 illustrates how the Census Bureau would handle ACS data for a city that annexed territory in April of 2006. In 2008, the Census Bureau will tabulate 1-year estimates for 2007 and 3-year estimates based on data from 2005, 2006, and 2007 using the new boundaries that were in effect on January 1, 2007, which would reflect this annexation. However, the Census Bureau will not update the 2005 and 2006 1-year estimates that were released previously and had used the old boundaries. Therefore, for the city in this example, the 2007 1-year estimates will be based on different boundaries than those used in 2005 and 2006 1-year estimates and the 3-year estimates released in 2008 would not be equivalent to an aggregation of the *published* 2005, 2006, and 2007 1-year estimates.

Table 5 points out an important feature of how the Census Bureau processes ACS data that is broader than the issue of boundary changes. In creating its 3-year and 5-year estimates, the Census Bureau combines the data from all the relevant years and calculates population, household, and household characteristics for the relevant geography. The data from previous years are retabulated together with the new data from the most recent year.

Table 5. **How a Boundary Change in 2006 Will Affect ACS Estimates**

ACS 1-year estimates	Year data collected	Year tabulated and released	Boundaries as of
2005 ACS estimates	2005	2006	January 2005
2006 ACS estimates	2006	2007	January 2006
2007 ACS estimates	2007	2008	January 2007
2005–2007 ACS estimates	2005, 2006, 2007	2008	January 2007

Source: U.S. Census Bureau.

[10] The most recent "Subject Definitions" can be found at <http://www.census.gov/acs/www/UseData/Def.htm>.

Rules Used to Determine Residency

An important technical difference between the ACS and the decennial census involves the rules used by the Census Bureau to determine where to consider a person a resident if that person lives or stays in different places at different times during a year. The Census Bureau refers to these rules as "residence rules." See Appendix 2 for more information about the census and ACS residence rules.

Household members who split their time between two or more housing units may be considered residents of different places in the ACS than they would have under the census residence rule. Thus, the ACS will base its characterization of the households and people in a given area on a slightly different "population" of households than either Census 2000 or the 2010 Census would use.[11] This difference in residency rules is unlikely to result in any measurable difference in the characterization of the local population except perhaps in highly seasonal areas where a sizable portion of the housing stock consists of second homes occupied for more than 2 consecutive months at a time. Resort communities in which families rent units for 1 or 2 weeks would not be affected by the difference in residence rules.

The second consequence involves the definition of a "vacant" unit. The ACS would classify fewer units as vacant in seasonal areas than the decennial census would. Users should look on this consequence as the result of a different concept of what constitutes vacancy rather than as a measurement error.

Sample Size and Statistical Precision

In 2005, there were approximately 124 million housing units in the United States.[12] The ACS annual sample of close to 3 million housing unit addresses represents an overall sampling rate of approximately 2.4 percent. The 3-year ACS estimates rely on a sample of approximately 7.2 percent and the 5-year ACS estimates rely on a sample of approximately 12 percent. Due to the use of subsampling in the final stage of data collection, the final sampling rates are lower than this. By contrast, the Census 2000 long form involved an overall sample of 16.7 percent, although the actual sampling rate varied by location based on population size, as does the ACS sample. In general, Census 2000 had a larger sample and therefore provided more precise estimates than should be anticipated from any of the ACS estimates.

To illustrate how the ACS compares with Census 2000 with respect to statistical accuracy, Table 6 contains an example using estimates of the number of noncitizens in five cities of various sizes. Recall that the margins of error reported in Table 6 represent the amount that a user would have to add to and subtract from the estimate to obtain a range about which the user could be 90 percent confident that the actual number of noncitizens would fall into that range. For example, a user would be 90 percent confident that the number of noncitizens in Dallas around April 1, 2000, was between 232,106 (234,829 – 2,723) and 237,552 (234,829 + 2,723).

Table 6. **Comparison of Estimated Margins of Error on Census 2000 Long-Form Estimates and ACS Estimates for Five Cities**

	2000 Census long form			2006 ACS		ACS margin of error		
	1	2	3	4	5	6	7	8
	Total 2000 population	Noncitizens	Margin of error (est.)	Total 2006 population	Noncitizens	1 year	3 year (est.)	5 year (est.)
Dallas, TX	1,188,204	234,829	2,723	1,192,538	259,182	11,894	6,867	5,319
Portland, OR	529,025	44,359	1,190	539,950	44,248	4,719	2,725	2,110
Newark, NJ	273,546	44,645	1,212	266,736	51,730	6,511	3,759	2,912
Gainesville, FL	95,605	5,488	451	105,714	6,461	1,656	961	745
Youngstown, OH	82,026	559	113	70,459	951	322	186	144

Source: U.S. Census Bureau.

[11] In this sentence, "population" is used in the statistical sense as the collection of items from which a sample is drawn for the purpose of describing that collection of items.

[12] *American Housing Survey for the United States: 2005*, Table 1-A1, issued August 2006.

The margins of error for Census 2000 noncitizen population counts in column 3 were estimated using procedures recommended by the Census Bureau.[13] The margins of error for the 2006 ACS 1-year data, reported in column 6, were obtained from the published ACS data. The estimated margins of error for the 3-year and 5-year ACS estimates in columns 7 and 8 were derived by assuming that the estimated counts of noncitizens would remain the same and that the samples would be 3 and 5 times larger respectively.[14]

Several important observations can be drawn from Table 6 that apply to most examples comparing data from Census 2000 and the ACS and within the ACS for 1-year, 3-year, and 5-year estimates:

- Census 2000 estimates have associated margins of error even though the Census Bureau did not publish them with the long-form data. In some cases, for example, Youngstown in Table 5, the margin of error is large relative to the estimate.
- As a general rule, margins of error are smaller for the larger areas. Looking at the long-form data, the margin of error is roughly 1 percent of the estimate for Dallas and 20 percent for Youngstown. Statistical accuracy depends upon the number of observations in the sample and not on the sampling rate. The long-form sampling rate of 12.8 percent for Dallas resulted in a sample of over 150,000, while the sampling rate of 15.5 percent for Youngstown produced a sample of less than 13,000.
- As expected, the ACS margins of error are larger than the corresponding margins of error associated with the long-form data because of the difference in sample size.
- The ACS margins of error are smallest for the 5-year estimates and largest for the 1-year estimates.

For Census 2000 data, most data users treated the long-form data as "counts" and did not consider sampling error. If one needed a count of noncitizens in Youngstown, one used 559 and ignored the possible margin of error. To correct this misconception, the Census Bureau has intentionally included margins of error with ACS data and much of its information for users has attempted to improve understanding of data reliability, quality, and associated measurement error. For many areas, ACS users will eventually be able to choose between 1-year, 3-year, and 5-year estimates. The 1-year estimates are the least precise but the most current, while the 5-estimates are the most precise but the least current. Users will have to balance currency against precision. The Census Bureau's decision to publish the associated margins of error will enable federal users to make this choice with a clear understanding of the precision of each of the choices.

Combining or Comparing ACS Estimates

Users will often want to combine or compare ACS estimates. With ACS data, users can easily add estimates,

Examples of Combinations and Comparisons Using ACS Data

(The numbers preceded by a B or C are the numbers of the detailed tables available on the American FactFinder—see Section 7.)

- *Combining two ACS estimates within the same table for the same geography*. From the table on educational attainment (C15002), one might want to combine the four rows containing estimates of males with bachelor's degrees and with graduate or professional degrees and females with bachelor's degrees and with graduate or professional degrees.
- *Combining ACS estimates from two different geographies*. One might want to estimate the number of owner-occupied homes valued at $1 million or more in DuPage and Lake Counties, Illinois, using Table C25075 for both counties.
- *Forming ratios or proportions using ACS estimates*. One could use the data in Table C18008 to calculate the proportion of the male population between 16 and 64 years old who have an employment disability.
- *Comparing ACS estimates from two different areas*. One might compare median household income in Denver, Colorado, in 2006 with median household income in Salt Lake City, Utah, in 2006 using Table B19013.
- *Comparing an ACS estimate for two different years or periods for the same geography*. One could compare median household income in Denver, Colorado, in 2005 with median household income in Denver, Colorado, in 2006 (Table B19013).

[13] *Accuracy of the Data: Census 2000 Summary File 3 (SF 3)—Sample Data*, Chapter 8, which can be found at <http://www.census.gov/prod/cen2000/doc/sf3.pdf#page=933>.

[14] Margins of error are inversely proportional to the square root of the sample sizes. The example is simplified because (a) the ACS 1-year, 3-year, and 5-year estimates would differ from one another as they would be based on different (if overlapping) samples and (b) because the sizes of the multiyear samples would not be exactly 3 and 5 times as large as the 1-year sample. Note that generalized variances were used to compute the margins of error for the Census 2000 sample estimates.

take ratios or proportions, or compare estimates from different places or time periods. However, correct data usage requires that users consider the statistical precision of the resulting combination or comparison. Fortunately, calculating the required measures of statistical precision using the information furnished in the ACS tables is straightforward, if somewhat complex. Appendixes 7 and 8 explain how to calculate the correct measures of precision for each of the five examples listed in the text box titled "Examples of Combinations and Comparisons Using ACS Data."

Nonsampling Error

The imprecision that results from sampling is just one of the reasons that a statistic can differ from the true value of what it is supposedly measuring. All the other reasons are grouped together under the term "nonsampling error." Nonsampling errors include incomplete coverage in sample design or execution, the inability to obtain any information for selected units (unit nonresponse), failure of respondents to answer specific questions (item nonresponse), wrong answers, and errors in processing the data.[15]

The ACS has some important advantages over Census 2000 with respect to nonsampling error. The ACS has a three stage data collection procedure: 1) self-response mail questionnaires, 2) computer-assisted telephone follow-up for nonresponse, and 3) computer-assisted personal visit interviews on a subsample of the remaining nonresponding sample addresses. Unlike the nonresponse follow-up operations conducted during Census 2000—operations that relied on temporary enumerators hired for the once-a-decade workload—the Census Bureau carries out ACS nonresponse data collection with a permanent telephone and field interviewing staff thoroughly familiar with the survey and its content.

Studies have indicated that the ACS procedures produce lower unit nonresponse and item nonresponse than experienced by the decennial long form.[16] While the impact of sampling variation can be easily translated into measures such as the margins of error in Table 5, it is difficult to actually measure how improvements in unit and item nonresponse translate into more accurate data. Survey experts generally believe that nonsampling error is the largest source of error in large-scale, modern surveys. [17]

[15] Appendix 6 discusses each of these sources of nonsampling errors.

[16] Susan P. Love and Deborah H. Griffin, *A Closer Look at the Quality of Small Area Estimates From the American Community Survey*, paper presented at the 2003 Joint Statistical Meetings in San Francisco, California, August 3–7, 2003.

[17] For example, on page D-1 of *American Housing Survey for the United States: 2005*, the Census Bureau concludes with respect to that survey: "Incomplete data and wrong answers are usually the largest source of errors, larger than sampling errors."

Sample Size and Suppression of Data

Rigorous provisions to protect the confidentiality of information supplied by respondents govern both the ACS and the decennial census. The Census Bureau has developed a number of safeguards to maintain confidentiality. Generally the safeguards applied to the ACS are the same as those applied to the long-form sample.[18] Because the application of a specific safeguard is usually triggered by the small number of observations used in a table or in a specific cell within a table, ACS tables are more likely to require safeguarding because its sample sizes are smaller.

An additional restriction is applied to ACS tables with 1-year estimates and 3-year estimates. Two versions of many of the detailed tables are produced, one with the full detail and a second with less detail—a "collapsed" version. Both are released as long as they pass a reliability standard. If the range of variation for the estimates in the detailed table is too high, the table is not released.[19] If a collapsed version is available that does meet the standard, the collapsed table will be released. However, if a collapsed version has not been specified, no version of the table will be available for that area. This safeguard was imposed to prevent users from using unreliable estimates; it was not used in tables from the 2000 long-form sample and will not be used in tables based on the 5-year ACS estimates. [20]

The safeguards listed above apply to published tables. The Census Bureau allows federal agencies to purchase customized tables but applies additional safeguards to these special products. The additional safeguards involve both minimal sample size rules and rounding rules.[21]

Public Use Microdata Sample (PUMS) files are also available. In PUMS, all personal identifiers and address information are stripped from the sample of individual ACS response records. Some variables are also coded to protect confidentiality. The geographic location of the respondents represented in these PUMS file records is no more precise than the defined Public Use Microdata Areas (PUMAs), geographical units containing an average of about 100,000 people.

[18] See *American Community Survey Design and Methodology Technical Paper* at <http://www.census.gov/acs/www/Downloads/tp67.pdf>.

[19] The ratio of the standard deviation to the mean is called the coefficient of variation. If the median coefficient of variation for a table exceeds 0.61, the Census Bureau will not publish the table.

[20] To see how this safeguard affects ACS data, the reader can compare Tables B25014 and C25014 (persons per room by tenure) for Atlanta, Georgia, and Columbia, South Carolina, in the 2006 ACS detailed tables. Table B25014 has 10 independent cells and is available only for Atlanta. Table C25014 collapses the 10 independent cells into 4 independent cells and is available for both Atlanta and Columbia.

[21] Tables by area generally require at least three cases and a mean cell sample size of three. The rounding rules are: 0 remains 0, 1–7 rounds to 4, and 8 or greater rounds to nearest multiple of 5.

Issues With Specific Variables

Disability. The Census 2000 estimates of disability, as well as estimates of disability in the ACS demonstrations prior to 2003, were unreliable. The Census Bureau redesigned this question in 2003 and the change produced a substantial downward shift in the percentage of people with disabilities between the Census 2000 long-form sample and the 2005 ACS. Comparisons between ACS years are not affected, but comparisons with the Census 2000 long-form sample and with the pre-2005 demonstrations of the ACS are not advised.[22]

Migration. As noted previously, the ACS asks a different question on mobility than the long form. The Census 2000 long form asked whether a resident lived in the same unit 5 years previous, while the ACS asks whether a resident lived in the same unit last year. Comparisons across ACS years are not affected but comparisons with the Census 2000 long form are not appropriate.

Vacancy. Because the ACS uses a "current residence" rule rather than the "usual residence" rule used by the decennial census, the ACS has a different concept of housing vacancy than the long form. This difference in concept makes it inappropriate to compare ACS vacancy rates with vacancy rates from the decennial census. However, the conceptual differences would result in only minor differences in the measurement of housing vacancy. More importantly, the ACS has a recognized downward bias in measuring vacancy because of the design of its data collection; this bias is unrelated to the differences in concept.[23]

Evaluations of the ACS

The Census Bureau tested the concept of continuous measurement for many years before the first full ACS in 2005, beginning with four sites in 1996. In 1999, the Census Bureau expanded its experimental ACS data collection to 31 sites (in 36 counties) across the country to test the questionnaire and data collection techniques. A nationwide sample of about 800,000 addresses—called the Census 2000 Supplementary Survey (C2SS)—was conducted in 2000, and the Census Bureau began releasing data for the nation, large states, and large areas in 2001.

The test sites and demonstration period allowed the Census Bureau to study the ACS concept in depth, and the analysis is available in a series of papers that can be found on the ACS Web site. In particular, the Census Bureau was able to make direct comparisons between data from the Census 2000 long form and ACS data from the combined 1999, 2000, and 2001 surveys conducted in the 36 ACS test areas. This evaluation compared measures of both sampling and nonsampling error between ACS and decennial census data down to the tract level.

CENSUS BUREAU EVALUATIONS OF THE ACS

Users who are interested in the Census Bureau's analyses of the strengths and weaknesses of the ACS should visit the ACS Web site, particularly the section on "About the Data." This section contains:

- A comparison of the 1999–2001 ACS data with 2000 decennial data in the 36 test counties, <http://www.census.gov/acs/www/AdvMeth/acs_census/index.htm>.
- Eleven evaluation reports, <http://www.census.gov/acs/www/AdvMeth/Reports.htm>.
- A number of other methodology papers, <http://www.census.gov/acs/www/admeth/papers/papers1.htm>.

[22] For a discussion of the changes in the disability question, see: *Disability Data From the American Community Survey: A Brief Examination of the Effects of a Question Redesign in 2003*, Sharon Stern and Matthew Brault, U.S. Census Bureau, Housing and Household Economic Statistics Division, January 28, 2005, available at <http://www.census.gov/acs/www/AdvMeth/Papers/Papers1.htm>.

[23] A unit in the ACS sample is not considered vacant until an interviewer visits the unit and determines it is unoccupied. The 3-month period between the mailing of the questionnaire and the in-person interview provides time for a vacant unit to become occupied prior to the interview. This procedure produces a downward bias in measuring vacancy.

6. Making the Transition to the ACS

Statutory Basis for Using ACS Data

This section identifies issues that federal users may experience as they begin to use ACS data. The section focuses on programmatic uses, such as the allocation of funds by formula, but some of the same issues may arise in agency use of ACS data for research, evaluation, or policy development.

Data analysts and policy makers in agencies should work with their legal offices to ensure that all relevant staff knows that the long-form sample has been replaced with the ACS. The Census Bureau considers the ACS to be a component of the decennial census; all the statutory language in Census Bureau legislation that applies to the decennial census also applies to the ACS. As part of the decennial census, respondents are required by law to respond to the ACS questionnaire, and the Census Bureau is bound to protect their responses in the same way that it protects responses to the long form. While the Census Bureau considers the ACS to be part of the decennial census, it is up to each agency to interpret the agency's legislation and to decide how ACS data should be used. While it is unlikely, it is not inconceivable that agencies may need to work with their congressional committees to adjust legislative language to encompass the characteristics of the ACS.

EXAMPLE OF A STATUTORY CONSTRAINT

The legislation for most active federal programs was drafted long before the ACS was developed and was written with the decennial census in mind, often without distinguishing between the 100 percent and sample data. In most cases, the language that was drafted with the decennial census in mind should not present any problems or create any limitations in using ACS data. But this will not always be true. For example, the Department of Housing and Urban Development (HUD) uses decennial census data to allocate funds among states and jurisdictions for its Community Development Block Grant (CDBG) program. The CDBG legislation mandates that HUD use data "referable to the same point or period in time." It would appear that HUD would have to use the ACS multiyear data for all recipients even through ACS 1-year data would be available for states that receive CDBG funds separately from municipalities. While HUD may prefer using the multiyear estimates in allocations to both states and municipalities, the legislation eliminates any alternatives when using ACS data.

Initial Choices in Using ACS Data for Federal Purposes

When to start using ACS data? "When" boils down to the choice between 1-year, 3-year, and 5-year estimates. If an agency's requirements include data for areas with populations below 20,000 (i.e., small jurisdictions, school districts, census tracts), then the agency will probably have to wait until 2010 when the first 5-year data become available. The 3-year estimates for areas with at least 20,000 inhabitants become available in 2008. Even if 1-year data would be appropriate for the planned use, an agency may elect to wait for 3-year or 5-year data to take advantage of the greater reliability offered by the multiyear estimates.

How often to "refresh" ACS data? This choice involves statistical, policy, and workload concerns.

- Agencies that plan to use ACS data in allocation formulas for program funds may decide to use 5-year estimates for their greater precision and to update the 5-year estimates every year so that the allocations change smoothly. While this approach ensures that formula allocations evolve slowly, it runs the risk of picking up turning points in local conditions slowly, 2 to 4 years after conditions change.
- Agencies that plan to use ACS data to determine the eligibility of entities, such as counties or cities, for participation in federal programs may decide to use 5-year data for their greater precision and to update the eligibility determination only when nonoverlapping 5-year estimates become available. Waiting for nonoverlapping estimates ensures that entities do not experience "on-again/off-again" determinations every time new ACS data become available.

The decision on how frequently to replace one set of ACS data with a newer set depends upon the unique features of a particular agency use. Maximizing precision and minimizing the impact on the entities with which the agency deals are important factors, but agencies may choose to emphasize other considerations such as currency.

Workload is an issue that federal agencies need to focus on early in the transition to ACS data. In the past, federal agencies received new long-form data only every 10 years and usually no special provision was made for processing the new data; staff simply worked around this once-a-decade spike in their workload. With the potential of having to process new data annually, agencies may have to add staff to handle the additional workload associated with the ACS.

How to prepare their institutional customers for ACS data? The ACS will be new to the entities, such as states, counties, cities, and planning consortiums, with which an agency normally deals. They may not be prepared for the changes that will accompany having new data every year. If agencies use formulas to allocate program funds, in the past the amounts received by an entity would vary from one year to the next only as the overall program budget changed—except for a once-a-decade change when new long-form sample data became available. With the ACS, an entity's allocation may change every year even if the program budget remains the same. Similarly an entity's eligibility to participate in a program may change more frequently than every decade.

Agencies will have to educate their institutional customers about the ACS and how it will affect the way the agency interacts with them. In addition, as the next two items will explain, agencies will have to prepare to deal with challenges from clients based on two features unique to the ACS, explicit recognition of statistical precision and multiple sets of estimates.

> **Statistical precision.** Consider the following situation: an agency receives new ACS data, processes it, and informs an entity that it is no longer eligible for program benefits based on the new data. The client challenges the agency's determination because the threshold for participation is within the published margin of error of the variable used to make the determination for that entity.
>
> **Multiple estimates.** An agency chooses to use 5-year ACS estimates in an allocation formula for program funds. One customer, New York City, claims that the 1-year estimates have sufficient precision, that the 5-year estimates have failed to pick up deterioration in its economy, and that using 1-year estimates would give it the higher funding that it deserves.

Agencies will not be able to rely on precedent to deal with these kinds of situations. The long-form sample produced only one set of estimates. Moreover, challenges based on statistical accuracy almost never occurred with past long-form sample data. The absence of published margins of error and the 10-year span between releases of new data created the sense that any measured change between censuses was a true reflection of actual changes. The publication of margins of errors and the short time period between data releases is likely to highlight questions of accuracy. Federal agencies should decide in advance how to handle problems like these and should consider incorporating how they use ACS data into formal rulemaking, which often provides the specific procedures for implementing statutory language. Going through the federal regulation process has several advantages. Because of requirements to publish proposed and final rulemaking in the Federal Register, rulemaking can help educate clients about the features of ACS data. It can solicit input on how ACS data are proposed to be used, input that may help agencies identify unexpected problems. Rulemaking can provide an agency with a stronger rationale for the procedures that it adopts.

Changes in the ACS

The ACS itself is going through some start-up changes that agency users should be aware of. These changes will have little or no effect on future ACS data but do affect the comparisons involving the first few years of data.

Data on group quarters. The ACS began collecting data on group quarters in 2006. The 2005 ACS population estimates include the household population only and are not comparable with total population estimates in the 2006 ACS. Data on group quarters are included in the 2006 1-year estimates and will be included in all future ACS 1-year estimates. The first 3-year estimates released in 2008 contain information on group quarters, as the Census Bureau adjusted the 2006 and 2007 group quarters data to represent all 3 years covered in the 3-year estimates released in 2008.

Household estimates. The Census Bureau changed the method of weighting for housing units and persons after the 2005 ACS resulting in a lower estimate of occupied housing units. That change impacts the estimate of households since the ACS estimate of households is equal to the estimate of occupied housing units. As a result, the household estimates are generally smaller in 2006 than in 2005. The same change causes the ACS estimate of total housing units to be slightly higher than the Census Bureau's official estimate of housing units in 2006. All future ACS releases will use the methods adopted in 2006. In the first 3-year estimate, released in 2008, the Census Bureau reweighted the 2005 data using the new method.

Examples of the Use of ACS Data by Federal Agencies

Federal agencies have already begun to use ACS data. Some agencies have been very proactive in making the transition to the ACS. For example, the Department of Transportation is working with its state and local partners to develop new specifications for the Census Transportation Planning Package (CTPP), a special tabulation of small area data to help local planning agencies assess their transportation needs. The CTPP was formerly based on the decennial long-form sample; the next tabulation will use the ACS. Here are two examples of how agencies are using ACS data:

- **Veteran Affairs (VA).** The VA Office of Policy and Preparedness is responsible for the Veteran Population Model that reports veteran population at the state level by count, period of service, age, and gender annually. Formerly this was done by taking veteran counts from the decennial census, adding discharges as reported by the Department of Defense, subtracting estimated deaths, and adjusting for migration. When new decennial long-form sample data became available, adjustments were made for the previous 10 years based on the new decennial census counts. The ACS provides estimates of the number of veterans at the state level by age and gender and by period of service. These estimates are based on a uniform methodology that does not require use of any data other than ACS data.

- **Housing and Urban Development.** Every year HUD estimates for every metropolitan area and every nonmetropolitan county the income limits that determine the eligibility of households for federal housing assistance. The Department of Agriculture, the Internal Revenue Service, and the Department of Veteran Affairs also use HUD's income limits. Formerly HUD derived income limits using a special tabulation of the long-form sample data and updated the estimates annually using wage data from the Bureau of Labor Statistics (BLS). Now HUD uses a weighted average of ACS and BLS data to update areas with populations of 65,000 or more. For areas with fewer than 65,000, HUD uses the ACS data to update at the state level and then uses BLS data to transform the state updates to smaller areas. When ACS releases multiyear estimates, HUD plans to make further changes in its procedures for calculating income limits.

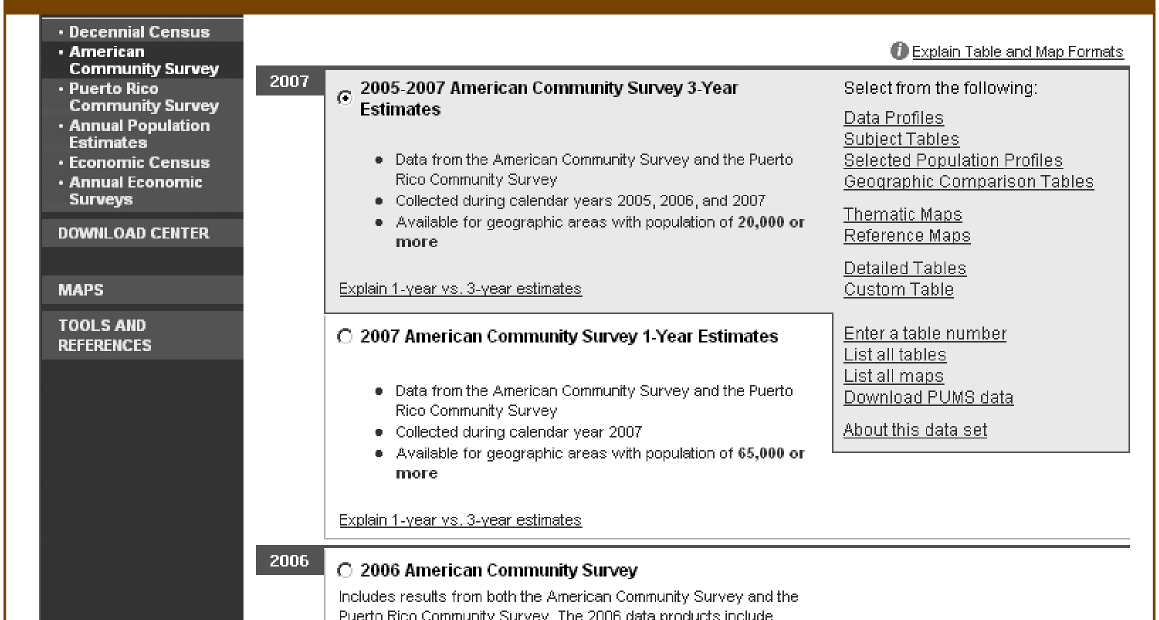
7. How to Obtain ACS Data

This section explains the most common ways in which federal users can access the ACS data that they will need to carry out agency business. This includes an overview of how to access tables via the American FactFinder, Summary Files by File Transfer Protocol, and Public Use Microdata Sample files. Custom tabulations are also discussed.

The American FactFinder (AFF)

The discussion starts with the American FactFinder (AFF) because this data dissemination tool is the simplest way to obtain specific ACS tables and because the AFF provides a convenient introduction to ACS products.

Figure 1. **Data Presentation Options Available on the American FactFinder**



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

One can access the AFF directly from the Census Bureau's home page, from the ACS home page, or simply by using the link for AFF at <http://factfinder.census.gov>. After accessing the AFF's home page, highlighting "data sets" on the left margin allows one to select decennial census data, ACS data, or another Census Bureau data series. Clicking on "American Community Survey" provides access to the most recently released ACS data, as well as data from earlier full implementation releases.

For example, clicking the radio button next to "2007 American Community Survey" reveals a series of options on the right-hand side of the Web page. (See Figure 1.) The first seven options are different types of prepared ACS products. Users should examine each of the products to see what types of tables are readily available. We will focus on the "detailed tables." Users should review the list of "detailed tables" because the detailed tables are the most thorough and complete tables published by the Census Bureau using ACS data; they are also the building blocks used to construct other products available on the AFF.

Clicking on "detailed tables" leads to a new page where the user can select the geography for the desired data. One can choose multiple geographies, including geographies of different types, such as, Texas, the Houston-Sugarland-Baytown metropolitan area, and Houston city. After selecting the appropriate geography, the AFF asks the user to select from among approximately 1,300 tables. The tables are organized by keywords to help users find specific topics. The user can select multiple tables. After the selection of both geographies and tables, the AFF displays the desired tables for the selected geographies. The tables can be printed or downloaded in various formats, including a Microsoft Excel workbook.

Summary Files

Federal users will most likely want a number of detailed tables for a large number of geographies. For example, an agency that uses ACS data to allocate funds at the state level would want data on each of the ACS variables used in the allocation formula for all states and the District of Columbia. The Census Bureau packages the detailed tables into a series of summary files that can be downloaded.

Users can find summary files at the Census Bureau's FTP (File Transfer Protocol) site, <http://www2.census.gov/acs2006/Final_Summaryfile/>. This site provides summary files for the United States and for each state and also includes documentation and instruction files. In particular, users should see the "Revised_ACS_2006_SF_Tech_Doc.pdf" file for information on how to read the summary files. The summary files are designed to be easily accessible through statistical programs such as SAS.

Custom Tabulations

Federal agencies can request and purchase custom tabulations of ACS data from the Census Bureau. Custom tabulations allow agencies to specify cross-tabulations of variables not available in the detailed tables or tabulations of data for unique geographies or for specialized subsets of the population. All custom tabulations must be approved by the Census Bureau's Disclosure Review Board to assure that the information conforms to the confidentiality requirements imposed by the statutes governing the decennial census. All data in custom tabulations are rounded and there are restrictions on the minimum number of sample cases for a given table and for each cell within a table. Custom tabulations are produced after the Census Bureau completes processing the regular products for each annual ACS data release cycle.

Agency staff interested in purchasing custom tabulations should go to <http://www.census.gov/acs/www/Products/spec_tabs/index.htm> or call 301-763-INFO to identify the ACS contact within the agency who will arrange for a meeting with appropriate staff at the Census Bureau.

Public Use Microdata Sample (PUMS) Files

All regular ACS products involve aggregations of responses from sampled housing units and group quarters within the geography being tabulated. The Census Bureau never releases completed individual questionnaires. It does, however, take a sample of interview records, remove all identifying information, and assign the response records to specially defined geographical areas called Public Use Microdata Areas (PUMAs). PUMAs are a set of contiguous census tracts that subdivide a state into a set of nonoverlapping areas each with a minimum population of 100,000. The Census Bureau releases the modified data in PUMS files. A complete description of the PUMS can be found at <http://www.census.gov/acs/www/Products/PUMS/>.

Federal users would most likely take advantage of the PUMS files for policy research purposes. The ability to look at individual records allows the user to construct cross-tabulations not available in the detailed tables and to perform regression analyses. The geographic areas are not often convenient for programmatic purposes.

Users can download the PUMS data into data processing programs such as SPSS or SAS and create their own tabulations, or they can use the Census Bureau's DataFerrett tool to analyze the PUMS data. Information about DataFerrett can be found at <http://dataferrett.census.gov/>.

8. Future of the ACS

ACS Data Become Available for All Geographies in 2010

Table 3 explains what ACS products—1-year estimates, 3-year estimates, and 5-year estimates—will be available for different geographies. The 1-year estimates were first released in 2006, the first 3-year estimates were released in 2008, and the first 5-year estimates will be released in 2010. Beginning in 2010, the Census Bureau will issue a new set of 1-year estimates, 3-year estimates, and 5-year estimates every year for geographies entitled to these estimates.

Future Evolution of the ACS at the Census Bureau

The long form evolved over time and so will the ACS. As society changes and as Congress adds or changes programmatic requirements, some questions will need to be reworded, other questions will need to be dropped, and new questions will need to be added. Users will demand new products or modifications to existing products. The Census Bureau anticipates that change will be needed and welcomes suggestions on products and content. As agencies gain experience working with ACS data, they should let the Census Bureau know what works well and where improvements could be made.

With respect to changes in content, a major content test was conducted in 2006 and a substantial revision of the questionnaire was implemented in January 2008. The most significant change was the addition of questions related to health insurance, marital history, and VA service-connected disability ratings. Agencies can always discuss additional changes or possible needs for new questions with the Census Bureau. However, change will be infrequent due to the complexity of implementing changes and the need for data to be consistent over a 5-year period.

Agencies should also take into consideration that the purpose of all ACS questions is to obtain information to carry out a legislative directive or critical programmatic need. As noted in Section 3, all the questions on the ACS questionnaire have been carefully vetted to ensure that they are fully justified and that the data are needed at small areas of geography that cannot be obtained through other federal surveys. Proposed new questions will receive the same review.

Future Evolution of the ACS Outside the Census Bureau

The ACS, like the decennial long form before it, is part of the nation's economic infrastructure. Federal agencies, state and local governments, businesses, and private citizens have used the output from past decennial long-form samples for a wide variety of purposes. Initially users will most likely plug ACS data into applications that formerly used long-form data but, as users become accustomed to receiving new data every year, they will undoubtedly find creative new applications for ACS data.

Once federal agencies understand the ACS and have mastered its features, they should contemplate how to take advantage of its unique characteristics. More frequent updating of program parameters and eligibility determinations are obvious possibilities. Other ideas include combining ACS data with data from smaller but more focused surveys to create efficient statistical models for small area estimation and using ACS responses to design highly targeted specialty surveys that can obtain important information at lower cost and with less burden on respondents. The ACS provides federal agencies with opportunities not available with the decennial long form; agencies should prepare to take advantage of these opportunities.

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 margin of error to the estimate (for the upper bound) and subtracting the margin of error from the estimate (for the lower bound). All published ACS margins of error 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.

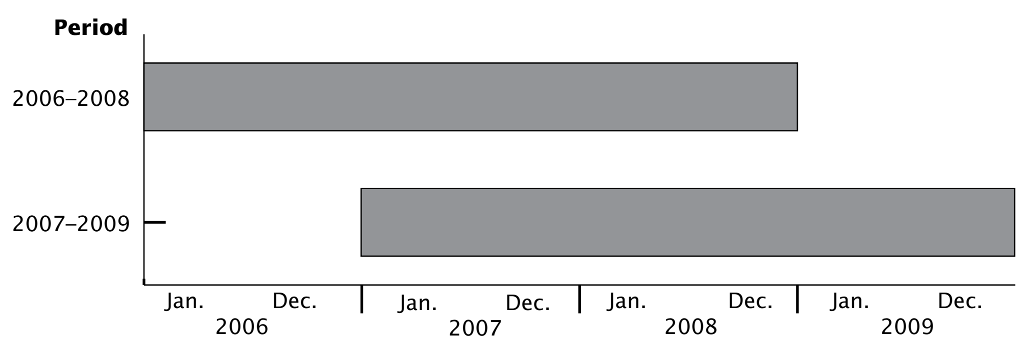
You certainly can compare characteristics for areas with populations of 30,000 to areas with populations of 100,000 but you should use the data set that they have in common. In this example you could use the 3-year or the 5-year estimates because they are available for areas of 30,000 and areas of 100,000.

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**

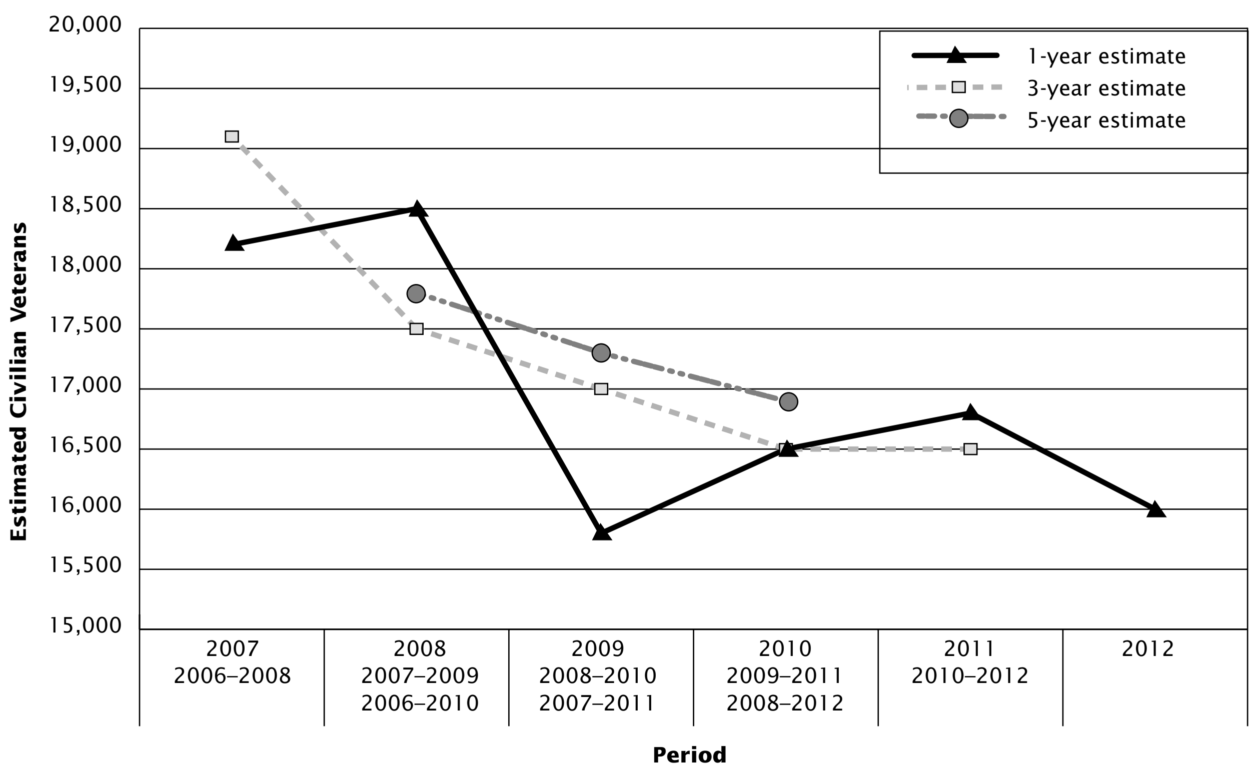


Source: U.S. Census Bureau.

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.

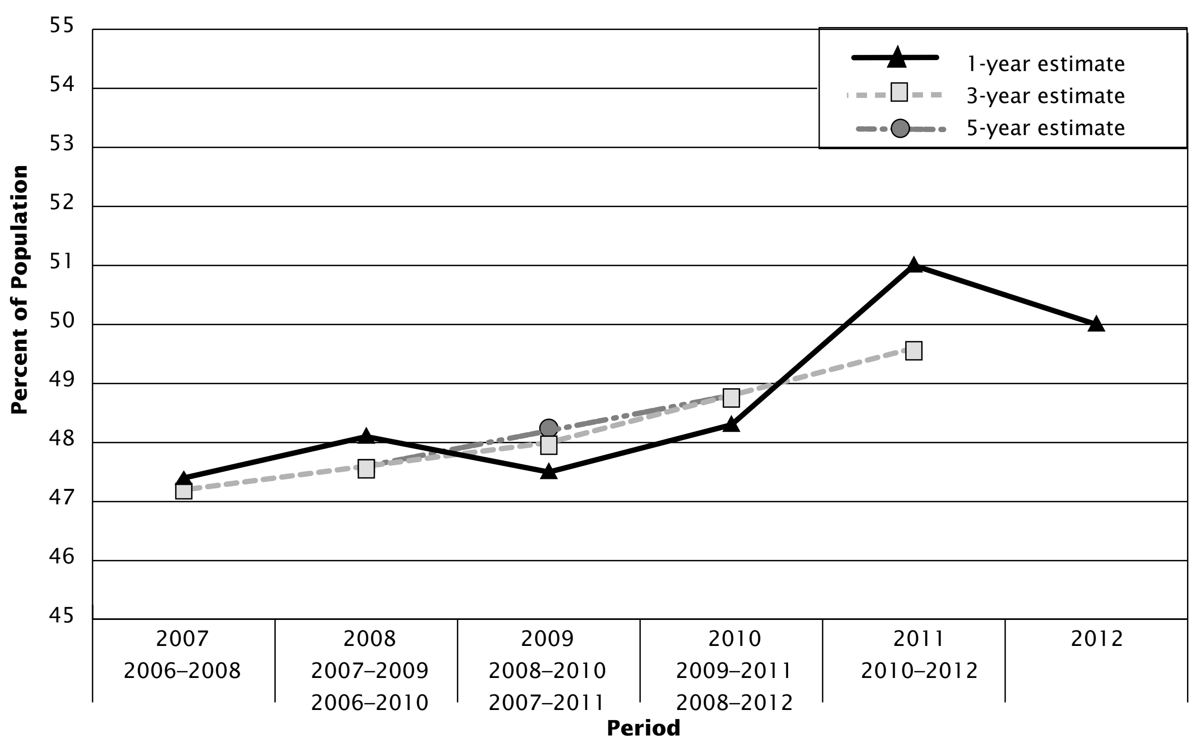
Local area variations may occur due to rapidly occurring changes. As discussed previously, multiyear estimates will tend to be insensitive to such changes when they first occur. Single-year estimates, if associated with sufficiently small CVs, can be very valuable in identifying and studying such phenomena. 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:

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

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.

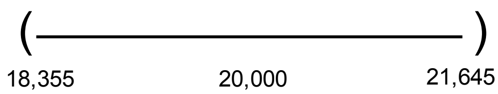
[3] Users are cautioned to consider logical boundaries when creating confidence intervals from the margins of error. For example, a small population estimate may have a calculated lower bound less than zero. A negative number of persons doesn’t make sense, so the lower bound should be set to zero instead.

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



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 c^{th} 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.

$\hat{p} = \frac{\hat{X}_{num}}{\hat{X}_{den}}$ 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).

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

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.

$\hat{R} = \frac{\hat{X}_{num}}{\hat{X}_{den}}$ 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.

$$MOE^2_{num} + (\hat{R}^2 * MOE^2_{den}) =$$

$$(989)^2 + [(0.7200)^2 * (1{,}328)^2] =$$

$$978{,}121 + 913{,}318.1 = 1{,}891{,}259.1$$

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:

$$MOE_{A\times B} = \pm\sqrt{A^2 \times MOE^2_B + B^2 \times MOE^2_A}$$

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

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

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.
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).

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

Algebraically, the significance test can be expressed as follows:

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).

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, 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.