A Compass for Understanding and Using American Community Survey Data

What Users of Data for Rural Areas Need to Know

Issued
August 2009

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

Paul R. Voss, Senior Spatial Analyst, Odum Institute for Research in Social Science, University of North Carolina at Chapel Hill, **Katherine J. Curtis**, Department of Rural Sociology, University of Wisconsin-Madison/Extension, and **Daniel Veroff**, Director, Applied Population Laboratory, Department of Rural Sociology, University of Wisconsin-Madison/Extension, 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.

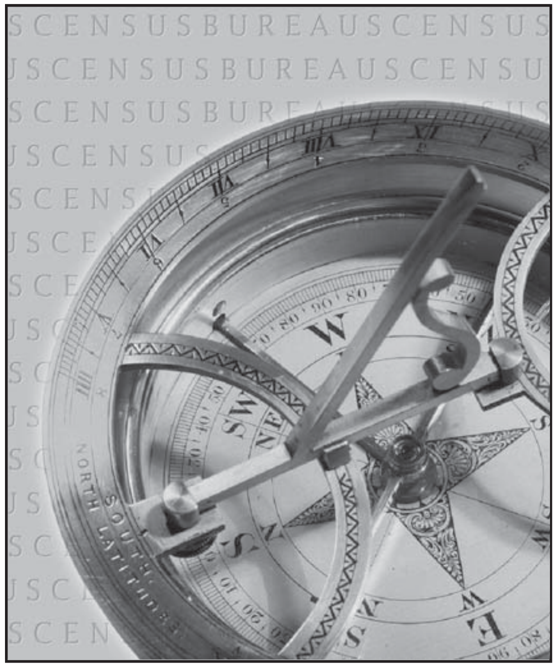
Other individuals who contributed to the review and release of these handbooks include **Dee Alexander**, **Herman Alvarado**, **Mark Asiala**, **Frank Ambrose**, **Maryam Asi**, **Arthur Bakis**, **Genora Barber**, **Michael Beaghen**, **Judy Belton**, **Lisa Blumerman**, **Scott Boggess**, **Ellen 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

Issued August 2009

What Users of Data for Rural Areas Need to Know



U.S. Department of Commerce
Gary Locke,
Secretary

Dennis F. Hightower,
Deputy Secretary

Economics and Statistics Administration
Rebecca M. Blank,
Under Secretary for Economic Affairs

U.S. CENSUS BUREAU
Robert M. Groves,
Director

Suggested Citation

U.S. Census Bureau,
A Compass for Understanding and Using American Community Survey Data: What Users of Data for Rural Areas Need to Know
U.S. Government Printing Office, Washington, DC, 2009.

ECONOMICS
AND STATISTICS
ADMINISTRATION

Economics and Statistics Administration

Rebecca M. Blank,
Under Secretary for Economic Affairs

U.S. CENSUS BUREAU

Robert M. Groves,
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.

Introduction

The purpose of this handbook is to introduce the American Community Survey (ACS) to people who care about, and use, social, economic, and housing data for rural communities. This handbook addresses ACS data issues for areas with populations as large as 20,000. That's a somewhat high-population threshold for a handbook focusing on small, rural communities, as it sweeps under our microscope thousands of cities and other incorporated places along with virtually all rural territory. Our inclination is to remain focused primarily on small, rural areas, and our examples are designed to address rural, more than small area, issues.

This handbook is written with language that should not appear overly technical and specialized for most potential ACS data users. We do presume, however, that someone using this handbook to navigate through the possibly unfamiliar waters of the ACS does have some minimal understanding of the mission and purpose of the U.S. Census Bureau and at least some limited experience accessing data from the Census Bureau's Web site. The reader should have some appreciation for data gathered using a sample survey including the statistical imprecision of survey data (often referred to as the "margin of error") that is occasionally published with sample-based estimates, most often with the outcome of public opinion polls.

We use the device of a single case study to illustrate a variety of aspects of this new national survey. While the case study involves a small rural community faced with the requirement of updating its social and economic profile as part of a larger land use planning exercise, many of the data issues confronted by the local official also apply to small urban settlements and small cities. We show how results from the ACS are both similar to, and different from, data formerly gathered in conjunction with decennial censuses using the long-form sample survey. Data once available for small governmental units and other small geographic entities from familiar census files called Summary File 3 (SF3) and Summary File 4 (SF4) have been supplanted by data from the ACS.

This change has been guided partly by demands from community leaders for data that are more current and relevant than traditional long-form sample results. By the early 1990s, it had become clear that, for many areas, census long-form data were outdated and irrelevant almost before they were released some 18 months following the actual census. The introduction of the ACS has also been guided by ongoing and intensive discussions between Census Bureau staff and data users so that ACS data products would meet the needs of community leaders, regardless of the size of their community. The development of the ACS has benefited from a decade-long testing and evaluation period to ensure that the transition from census long-form data to ACS data is as smooth as possible and that training materials would be available to ease that transition. This handbook is one of several similar publications designed to meet this goal.

Data users who are not yet familiar with the ACS will be comforted by the similarities between the long-form survey and the ACS (e.g., similar questionnaire format and content and similar data products). However, for all ACS data users, there will be a necessary reorientation that alerts the data analyst to some major differences between the census long-form survey and the ACS. For example, the sampling strategy and rules for determining residence in the ACS are different from the long-form survey. These differences will require users of small area data to invest time and effort to gain familiarity with the new ACS and to discover how ACS data can be used in fresh and innovative ways to address questions frequently asked by rural community leaders, data analysts, and planners. The Census Bureau has provided several documents to assist in understanding these differences. Many can be found on the Internet at <http://www.census.gov/acs/www/>.

Technical and statistical issues related to using ACS data are presented as appendix material. We point to these technical materials, as necessary, when describing how one small community, the hypothetical town of Wolf Lake, confronted the need to consult ACS data to develop a comprehensive town planning document. In addition, the Census Bureau's Web site <www.census.gov> has a wealth of information regarding the ACS. A particularly useful link provides specific guidance regarding cautions when comparing ACS data with data from other data sources, including the Census 2000 long-form sample data at <http://www.census.gov/acs/www/UseData/compACS.htm>.

The American Community Survey

The ACS is an innovative national initiative designed to collect statistically reliable and timely information needed for critical government functions. See the "What Is the ACS?" text box for details. While the ACS was tested and evaluated in a handful of counties for more than a decade and nationally for 5 years, it only commenced full implementation in 2005. Counties and cities with populations of 20,000 and more have already received ACS data. Smaller geographic areas, specifically those with populations below 20,000, must wait until 2010 before ACS data are available for their communities. Thus, beginning in 2010, the ACS will provide data users across the country with annually updated information about America's small communities, neighborhoods, and rural areas.

The principal feature of the ACS that makes it different from the census long-form sample survey is that the ACS is a continuous, ongoing data collection effort. Every month, a fresh sample of about 250,000 addresses is identified for interviewing. The data products from the ACS resemble those produced from the decennial census. In contrast with traditional census products that approximated data for a point in time called Census Day, the tables of data from the ACS represent the accumulation of data gathered over the course of 1-year, 3-year, and 5-year time periods. Like the census long-form sample, a larger percentage of addresses are selected in sparsely settled rural areas. Despite this "rural oversample," ACS interviews must be accumulated over several years to achieve reasonable statistical reliability in smaller areas.

Table 1 shows the kinds and proportions of geographic areas for which cumulations of 12 months of data are sufficient to generate statistically reliable data. These cumulations generally are referred to as "1-year estimates" and are released by the Census Bureau every

What Is the ACS?

The American Community Survey (ACS) is a nationwide survey designed to provide communities with an updated statistical portrait every year. The U.S. Census Bureau, under the authority of Title 13, U.S. Code, Sections 141 and 193, conducts the survey. Title 13 also requires that the Census Bureau use this information only for statistical purposes. Thus, all statistical tables and public use files based on ACS results strictly maintain the confidentiality of individual responses.

Survey questionnaires are sent to approximately 250,000 addresses across the country every month. Addresses from which a questionnaire is not returned by mail are followed-up, first in an attempt to obtain the information by telephone, and then, for a sample of nonresponding households, in person by a Census Bureau field interviewer.

Based on responses from the series of 12 independent monthly samples each calendar year, the ACS can provide estimates of demographic, housing, social, and economic characteristics for all states, as well as for cities, counties, metropolitan areas, and population groups of 65,000 or more. These estimates, based on a full year's worth of collected ACS data, are called "1-year estimates." For less populated areas, such as rural villages and towns, 3 or 5 years of ACS data are accumulated to produce statistically reliable estimates of population and housing characteristics. Estimates for areas with populations of 20,000 or more are based on data collected over 3 years ("3-year estimates").

For rural areas, urban neighborhoods, census tracts, block groups, and population groups of fewer than 20,000 people, it will take 5 years to accumulate enough survey data to achieve data estimates with statistical reliability that is similar to that of the Census 2000 long-form sample survey. These latter survey estimates, called "5-year estimates," will be published for areas with small populations each year beginning in 2010. The Census Bureau will also produce 3-year and 1-year data products for larger geographic areas.

A detailed description of ACS data collection methodology and the survey's sample design may be found at <http://www.census.gov/acs/www/SBasics/desgn_meth.htm>.

Table 1. **Major Geographic Areas and Type of ACS Estimates Published**

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.

year for geographic areas with at least 65,000 people. These areas will also receive 3-year and 5-year estimates. Also shown is the proportion of major types of geographic areas for which cumulations of 3 years or 36 months of data ("3-year estimates") and cumulations of 5 years or 60 months of data ("5-year estimates") will be released. Table 1 also shows that of the 3,141 counties and county equivalents, 25 percent are large enough to receive 1-year, 3-year, and 5-year estimates every year. Most counties (and equivalents) will only receive 5-year estimates each year since they have populations of less than 20,000. All census tracts, ZIP Code Tabulation Areas, and census block groups, regardless of size, receive only 5-year estimates. Most cities, towns, census designated places, townships, and villages (minor civil divisions) will only receive 5-year estimates.

Table 2. **ACS Data Collection and Release Dates**

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

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

Table 2 summarizes the data release schedule for each of these types of estimates. For example, in 2011, when basic demographic data from the 2010 Census begin to be released, detailed social, economic, and housing data from the 5 years of ACS interviews spanning the period January 2006 to December 2010 will also be released. These releases will include the 1-year estimates, based on data collected in 2010 for areas with at least a 65,000 population; 3-year estimates, based on data collected in 2008 through 2010, for areas with at least a 20,000 population; and 5-year estimates for all areas.

Accessing American Community Survey Data

ACS data are presented in a number of different products, some of which are mentioned in the text box "ACS Data Products." They range from profiles of broad social, economic, housing, and demographic characteristics to very detailed tables that focus on a specific variable. ACS data products include thematic maps and tables that compare data across geographies.

These products cover a wide array of subjects, listed in Table 3 and are generally equivalent to the subjects released based on the long-form data from Census 2000. A highly useful link on the Census Bureau's ACS Web site provides a technical definition of each published data item in the ACS at <http://www.census.gov/acs/www/UseData/Def.htm>.

The ACS data products can be accessed through the Census Bureau's American FactFinder Web site: <http://factfinder.census.gov/home/saff/main.html>. From this main page, two links under the "Getting Detailed Data" section lead to the ACS. The "learn more" link provides users with a basic overview of the ACS and its uses, and a link to the Census Bureau's ACS home page. The "get data" link leads to the ACS data sets Web page where users can select the products and the subjects that fit their data needs. A world of alternative data options awaits the user here. For example, a mouse click on "detailed tables" opens a dialogue section that will appear familiar to anyone who has used the American FactFinder to obtain decennial census tables, except that here the tables are generated from ACS data.

ACS Data Products

While ACS data are currently being released each year for geographic areas with populations of at least 20,000, the survey is preparing for cumulations of data for the smallest geographic areas. The first 3-year estimates were released in 2008, based on data collected in 2005, 2006, and 2007. Beginning in 2010, based on data gathered from 2005 through 2009, annual 5-year estimates from ACS surveys will be released for all census geographic units, regardless of size, down to the census block group level of geography. These annual publications will include the following product types in addition to others:

- Detailed tables
- Subject tables
- Geographic comparison tables
- Thematic maps
- Data profiles

The Census Bureau also releases each year single- and multiyear public use microdata sample (PUMS) files based on a sample of housing units and people living in group quarters. These PUMS files permit the creation of customized tables and the detailed analysis of statistically "rare" population groups (e.g., households that include members spanning three generations). The microdata records from the PUMS files preserve the confidentiality of individual responses in a variety of ways. Tabulations of microdata can only be produced for states and predesignated Public Use Microdata Areas (PUMAs), each having a population of at least 100,000 people. This limitation of geographic detail helps ensure that PUMS data remain confidential.

The Census Bureau also plans to release two specialized data files twice each decade, beginning in 2012. These will be based on 5-year estimates:

- American Indian and Alaska Native Summary File
- Summary File 4 (detailed tables by race, ethnicity, and selected ancestry groups)

Table 3: **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: U.S. Census Bureau.

Getting ACS data from the American FactFinder generally involves a five-step process:

1. Select the ACS data set (for example, the 2007 ACS 1-year estimates or the 2005–2007 ACS 3-year estimates).
2. Select the data product from the options provided in the right-hand margin.
3. Select the geography (or geographies) of interest from the drop-down lists of available geographies.
4. Select the specific table or tables desired.
5. View or download the data. The results are provided on screen and can be downloaded in a variety of file formats (e.g., comma delimited, tab delimited, Excel).

More experienced data users may choose to download ACS data using the File Transfer Protocol (FTP). This option is provided on the ACS data release page at <http://www.census.gov/acs/www/Products/index.html> or directly at <http://www.census.gov/acs/www/Special/acsftp.html>.

For someone approaching the Census Bureau's ACS site and the American FactFinder interface for the first time, the process may appear intimidating. But a bit of practice retrieving data tables for different kinds of geographic units and for different years will fairly quickly make even a novice user comfortable. The best source for assistance in many cases will be the specialists at their State Data Centers: <http://www.census.gov/sdc/www/>. In the text box, "Examples Where ACS Data Might Be Used," we briefly list several examples.

Examples Where ACS Data Might Be Used

- A consortium of human services agencies and a rural county health facility use ACS data on age, poverty, disability, and access to transportation to examine issues of health, aging, and rural poverty. They use ACS data to compete for a variety of human services grants and loans and to tell the local story to pursue funding for a satellite clinic to serve low-income families and the elderly.
- A county public health department uses an array of ACS data (age, race/ethnicity, poverty, education level of parents, and single parents) as key indicators of populations at risk for diabetes and asthma in children. They supplement the ACS data with data from local clinics to help target education and intervention programs through community organizations and school systems. The goal is to teach children and their parents about healthy lifestyles, living conditions, and diets.
- A state low-income housing coalition uses PUMS data from the ACS on income by household size and contract rents to evaluate the supply of affordable rental housing.
- A school district compares ACS data with Census 2000 to track trends in the Spanish-speaking population. Combined with administrative data collected by the district, the ACS data help the district to reach a decision about the need to hire more bilingual teachers.
- In preparing an application for Community Development Block Grant funding for a new library/senior community center, a village uses ACS data to document that at least 51 percent of those benefiting from the project will be low-income or moderate-income individuals. The village also uses ACS data on age and income to make the case for a seniors' facility.

Using This Case Study

The following case study uses a fictitious e-mail dialog between a data user and a State Data Center demographer to touch on key issues and instructions about the use of ACS data. You will note that in addition to this dialog, summaries of important points are provided in the left margin. These summaries can serve as indices for finding guidance on particular topics. The data from the 1990 census, Census 2000, and the Population Estimates Program that are cited in this case study are real, although we have renamed the area Wolf Lake.

Data from the 2001–2005 ACS Multiyear Estimates Study were used in this case study as the 2005–2009 ACS estimates.[1] The 2005–2009 ACS margins of error were not based on the Multiyear Estimates Study; rather, they were approximated by using a constant relationship between the Census 2000 margins of error and the ACS margins of error. To illustrate certain concepts more clearly, reasonable (but fictitious) data are used for the 2010 Census and for several projections.

Introducing Wolf Lake

Located in the heart of Lake County, Wolf Lake is a town steeped in the folklore of mythical lumberjack Paul Bunyan. Local history and geography are a rich mixture of colorful legends of lumber camp high jinks, true stories of devastating forest fires and selfless heroism, and ever-embellished tales of invasive bears, record deep snowdrifts, and bigger-than-life game fish. The local chamber of commerce advertises Wolf Lake as a “vacation paradise.” Wolf Lake is a rural town of friendly people and 36 square miles of a welcoming blend of forests, lakes, rivers, and natural amenities that draw both tourists and new residents to the area.

The town is a functioning governmental unit with a town chairman, a three-member board of supervisors, and a single, paid, town employee—part-time clerk/treasurer, Pat Smith. A few relevant demographic statistics for the town are shown in Table 4.

Wolf Lake’s Census 2000 population of 958 showed an increase of almost 10 percent over the 1990 census population. The state demographic agency’s annual population estimates for 2007 imply an increase of about 35 people, representing a growth rate of almost 4 percent since 2000. An annexation of land from a neighboring town in 2006 accounts for some of that growth. A total of 44 housing units were added to Wolf Lake. Most of the town population resides in housing units. A single nursing home was added to the community since 2000. The town has nearly 1,500 housing units, but more than 70 percent of these generally are found to be vacant at census time (April). Many of these vacant units are large lakeshore homes owned by families from the metropolitan city roughly 160 miles to the south. These homes are capable of supporting year-round occupancy, and some are used for winter vacations when owners come up for brief holidays of cross-country skiing and snowmobiling. Mostly, however, these places are occupied for various intervals of time during the summer months when families come to Wolf Lake to swim and fish, and enjoy some of the other tourist-oriented recreations that spring to life in early May. Some of these families are “snowbirds” that spend the winter months in Sun Belt communities but return to Wolf Lake for the summer each year. In these summer months, the area also experiences a large increase in the population who work in these tourist areas—high school and college students in particular. A map of the town is shown in Figure 1.

Table 4. **Demographic Overview: Town of Wolf Lake**

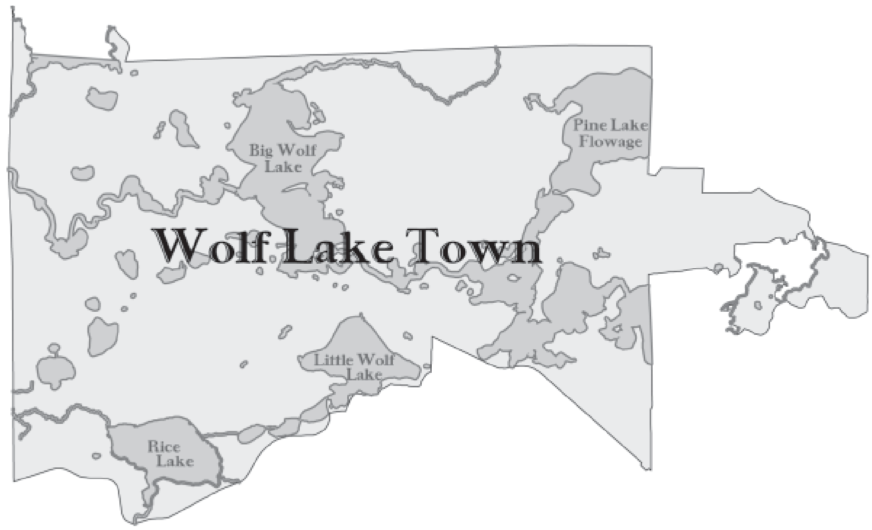
	1990	**2000**	**2007**
Population	884	958	993
In Households	884	958	NA
In Group Quarters	0	0	NA
Housing Units	1,240	1,407	1,461
Occupied	376	445	NA
Vacant	864	962	NA
Percent vacant	69.7	68.4	NA

NA Not available.

Note to the reader: Fasten your seatbelt. For this case study to be realistic, we must fast-forward to the fall of 2011. The ACS has been in full swing for over 5 years and the 2010 Census (with no long-form supplement) was conducted without major difficulty or controversy.

[1] The ACS Multiyear Estimates Study includes a series of single- and multiyear estimates for a sample of areas across the country. These estimates were produced to dress rehearse production systems and to provide a preview of multiyear estimates. For more information, see <http://www.census.gov/acs/www/AdvMeth/Multi_Year_Estimates/overview.html>.

Figure 1. **Map of the Town of Wolf Lake**



Source: U.S. Census Bureau.

The 2010 Census population count for Wolf Lake was 1,006—an increase of 48 persons (5 percent) over the Census 2000 count. The 2010 Census included 32 residents of the nursing home and a housing unit count of 1,474, with 1,022 housing units (69.3 percent) enumerated as vacant (see Table 5).

Table 5. **Demographic Change 1990 to 2010: Town of Wolf Lake**

	1990	2000	2010
Population	884	958	1,006
In Households	884	958	974
In Group Quarters	0	0	32
Housing Units	1,240	1,407	1,474
Occupied	376	445	452
Vacant	864	962	1,022
Percent vacant	69.7	68.4	69.3

NA Not available.

The Need for a Comprehensive Plan

In October 2011, the Lake County board of supervisors, upon advice from the county attorney, voted to bring the county into conformity with the State's Smart Growth Law. The decision meant that Wolf Lake would join other governmental units in the county and commence a comprehensive planning process. The county board offered the assistance of the county geographic information system officer to each town government, but the supervisors in Wolf Lake decided also to contact State Data Center demographers at the state university. Terry Adams, a university demographer, agreed to work with town clerk, Pat Smith, to assist in understanding, accessing, and using ACS data that are available on the Census Bureau's Web site at <www.census.gov>. They both agreed this would help speed the completion of the demographic, social, economic, and housing sections of the plan, and provide two sets of eyes to ensure that the data were extracted and interpreted properly. These four sections of the comprehensive plan cover a broad range of data and are intended to provide both a snapshot of current conditions and relevant trends as a backdrop for the town's new land use plan. Over the last 10 years or so, the town has been growing modestly but has been dealing with an aging population, loss of jobs from a 2008 manufacturing plant closing, and continuation of a long-familiar pattern: the exodus of young people as soon as they graduate from high school and, with troubling frequency, sometimes before graduating. So, along with the "snapshot" demographic data from the 2010 Census, the clerk was instructed by the town supervisors to locate other data that might address some of these issues.

Commence the Dialogue

Clerk Smith learns from the university demographer that she can access the data she needs on the Census Bureau's Web site. She is asked to find the link to the ACS where she'll be able to access rich, detailed socioeconomic data about Wolf Lake. She is encouraged by the demographer to read the documentation for the ACS before using the data. After briefly reviewing some of the documentation, she begins a series of e-mail exchanges with the university demographer, Terry Adams. Follow along.

Understanding ACS data requires preparation. Use the Census Bureau's Web site.

[November 4] Clerk Smith: Hi, Terry, I've spent much of the week reading some of the ACS documentation on the Census Bureau's Web site. I especially looked carefully at the ACS main page <http://www.census.gov/acs/www/>. It's all pretty clear, but there's really a lot there! I won't get to it all, and I'm not even sure I entirely understand some of what I've looked at, but I think I'm on board. It seems that I will only have 5-year estimates for Wolf Lake, right? The ACS data released a couple weeks ago cover the period 2006 to 2010. Can you tell me what that means, exactly? Is it as simple as averaging all 5 years' worth of data? Please remember, I'm not a statistician.

Most geographic areas in the United States will have only the 5-year estimates from the ACS to update their social, economic, and housing characteristics. But they will receive a fresh update every year.

[November 4] Demographer Adams: Great, Pat. I'm glad you found the documentation helpful. It sounds like we're ready to get to work. Let me take your questions one at a time. You are correct that you will use the 5-year estimates from the ACS. Of more than 25,000 incorporated places in the United States, about 92 percent will only have ACS data from the 5-year estimates. The overwhelming majority of governmental units across the country have populations below 20,000, which means that you're in good company. All of the other towns in Lake County will be using the 5-year estimates for their comprehensive plans, too. Lake County, as an entity, is large enough to have 3-year ACS estimates released for it, so the county has 3-year estimates for 2005–2007, 2006–2008, 2007–2009, and 2008–2010. The county will also have 5-year estimates for 2005–2009 and 2006–2010, and it is these latter two ACS data sets that would provide the appropriate county data for the towns in Lake County to use when comparing their trends with those of the county. These are the only two sets of ACS estimates currently available from which you'll be able to generate data for Wolf Lake showing detailed population and housing characteristics. But remember, these data are updated with fresh 5-year estimates every year.

When you are comparing ACS data for two different areas, be sure to compare the same types of estimates (e.g., 5–year to 5–year).

[November 7] Clerk Smith: And what about my second question? What do these 5-year estimates really mean? Are they some kind of average?

ACS period estimates may be interpreted as if they were averages of survey results gathered over the data collection period.

[November 8] Demographer Adams: I think you're asking how you should interpret a 5-year estimate, and you're on the right track. In many respects, the period estimates of social and economic characteristics based on 5 successive years of ACS interviews are pretty straightforward. They're more or less averages based on observations over the past 5 years, but the estimates are not simple averages determined by adding up all the data collected over the 5 years and dividing by 5. The way the Census Bureau calculates a period estimate is a little more complicated, but, in reality, it's not inappropriate to think of such numbers as being like an average. The estimates will reduce the high and low sample observations over the 5-year period. There are some issues that complicate matters just a bit, however, and I'll raise these when we discuss the differences between the census long-form sample and the ACS.

[November 8] Clerk Smith: Okay, I think I understand what these 5-year ACS estimates represent. I have used the census long-form data for these planning exercises in the past. Since you raised the issue, how do the 5-year period estimates from the ACS compare with what might have been discovered had the 2010 Census actually sampled and interviewed people using a long-form questionnaire?

ACS period estimates take the place of the traditional census long-form sample estimates.

[November 9] Demographer Adams: I suspect you're not the only town clerk asking this question about now, and I confess that it's taken me some time to understand the differences. The switch to gathering and publishing detailed social and economic characteristics based on the ACS represents a remarkable shift for data analysts in the kinds and amounts of data available. For small communities like Wolf Lake and for rural areas in general, the chief benefit of the ACS is that we're going to get updated socioeconomic and housing data every year. No more waiting for the next census plus a couple more years to get the data checked and published. But you need to be prepared to understand this new alternative to the long-form estimates. In some ways, this new survey is very different from past census samples, and there are important tradeoffs involved in this change that we all need to understand. Indeed, some of the challenges of the ACS are considerably more troublesome for small communities and rural areas than they are for large cities. Therefore, an understanding of these differences is going to be the key to the successful use of this new survey.

The ACS and the old long-form survey are different in several respects.

[November 9] Clerk Smith: Can you be specific about those differences?

The ACS is based on a continuous data collection strategy, and the sample results must be accumulated over time until the data can be reported with statistical reliability.

[November 10] Demographer Adams: You bet. The transition from traditional census long-form results to ACS results for most data analysts will be pretty straightforward, but for some it may not be quite so easy. You're already aware of one of the important differences. The 2010 Census collected data over a relatively short time period—March through June of 2010. In comparison, the ACS collects data continuously throughout the entire year. This continuous measurement nature of the ACS is certainly one difference that will require a new way of interpreting some of the ACS data. The ACS data are estimates based on interviews spanning a period of time. For Wolf Lake, 5 years of data are required before estimates with good statistical reliability can be released.

The way the census and the ACS determine whose characteristics are to be reflected in the estimates for a place differs.

In the census, most people are assigned to a "usual" place of residence.

In the ACS, people are assigned to a "current" place of residence, which may or may not be their "usual" residence.

Another difference is that the rule for determining who should be included in the ACS is different from the rule in the census. In the 2010 Census, households are enumerated, and their characteristics are associated with their "usual place of residence." This may or may not be where they happened to be living or staying on Census Day (April 1 in recent censuses) which makes collecting these data often difficult. However, households receiving the ACS questionnaire where they are living or staying will be interviewed, and their characteristics will be associated with that place, if their length of stay there is intended to be more than 2 consecutive months. The ACS refers to this concept as "current residence," and for people who meet this "2-month rule," the place where they are interviewed for ACS may or may not be their usual residence. It's possible that the ACS is including, as it should, some of the summer residents of Wolf Lake who may not have been included there under the census rules. Given that you know the Wolf Lake community better than I do, I encourage you to think about how the presence of part-time summer residents in Wolf Lake might make the ACS estimates different from the Census 2000 long-form results and get back to me with your thoughts.

[November 14] Clerk Smith: Let me be sure I understand this. Are you saying that people who vacation at Wolf Lake on weekends throughout the summer will be included in the ACS? I know they weren't included in those vacation homes in the census—they were vacant units.

[November 14] Demographer Adams: No. The ACS would treat those weekenders just like the census. It's the households who live there for more than 2 months that are interviewed as Wolf Lake residents by ACS. For example, people who spend the winter in warmer places but move back to Wolf Lake for the summer should be included in the ACS if their address was in sample in the summer. If that address was in sample in the winter, when they were living in that "warmer place," the unit should be classified like the census, as vacant. Also, people who work for the entire summer in Wolf Lake, but live somewhere else for the rest of the year, could be included in the ACS, but not in the census.

[November 14] Clerk Smith: Terry, I get it now and you're absolutely right. The data do look different on several measures. I looked at the 2005–2009 ACS 5-year estimates and compared them with the data from Census 2000 and from the 2010 Census. Here's some of what I see:

Table A. **A Few Wolf Lake Characteristics**

	Census 2000	2005–2009 ACS	2010 Census
Total housing units	1,407	1,587	1,474
Occupied housing units	445	624	452
Vacant housing units	962	963	1,022
Percent vacant	68.4	60.7	69.3
Owner-occupied housing units	358	405	361
Total population	958	1,545	1,006
Total civilian employed population	335	573	NA
Employed in retail trade	70	127	NA
Employed in arts, entertainment, recreation, accommodation, and food service	89	399	NA
Median age	49.6	39.5	49.2
Percent age 65 and over	25.7	22.3	24.3

NA Not Available.

The ACS data I found on the Census Bureau's Web site also contained a plus-minus margin of error for each of the estimates, but I need your help in interpreting these. I just wasn't sure what to do with them. Any advice?

[November 14] Demographer Adams: If it's all right with you, I'd prefer to not talk about the margins of error yet. They're very important, and we will discuss them, I promise. I like your data comparisons, and it does seem that the ACS is reflecting the characteristics of some of your summer residents. The ACS estimate of vacant units is less than the census count, and I think that's to be expected. Due to your increased population in summer months, some of the homes reported as "occupied" by the ACS apparently were considered to be vacant by the census. From what you've told me about the folks who spend their winters in states like Florida or Arizona and summers in Wolf Lake, I'm surprised to see an ACS lower median age and a lower ACS estimate of people who are 65 and over.

If a large portion of the households in an area lives there for several months, but not for most of the year, the community may find noticeable differences between census data and ACS data.

[November 14] Clerk Smith: Let me summarize what I think you were telling me about the differences that I might see in the ACS data that could be due to including some of the seasonal residents. In Wolf Lake, we do have some retired couples that live at Wolf Lake for only about 4 months but we also have younger people who come to work at Wolf Lake. I assume that the ACS data include both of these groups and that's why the median age in the ACS is a little lower. This seems to be supported by the increase in the percent of people employed in retail trade, arts, and entertainment.

[November 14] Demographer Adams: That sounds reasonable.

The margins of error (MOE) attached to every ACS estimate are very important and signal the extent of statistical uncertainty associated with the estimate. This uncertainty results from the fact that the estimate is based on a sample and not on a complete count.

Census long-form sample estimates also have levels of statistical uncertainty resulting from sampling error.

[November 15] Clerk Smith: Can we begin the conversation about the margins of error attached to each number? I know a little about MOEs. I know that they are related to a specific confidence level. I think I read that the Census Bureau uses a 90-percent level of confidence. Also, the MOEs are calculated from standard errors using this confidence level. MOEs give us an idea of how reliable, or precise, the estimates are.

[November 18] Demographer Adams: Yes, that's right. I'm glad you know the basics. There are a couple of things I should say about sampling error. First is something I wanted to mention this past Monday when you presented your Census 2000 data and compared those estimates with ACS 5-year estimates. Remember that the numbers you are familiar with from the Census 2000 SF3 file are also based on a sample. Most numbers in that file also have margins of error, and many of us simply ignored the fact that sampling error was associated with them because the MOEs were not published along with the long-form sample estimates. So, in addition to the MOEs for the ACS estimates, you also need to consider the MOEs for the long-form estimates from Census 2000. The Census Bureau provides very useful technical documentation for SF3 that shows how to compute estimates of these MOEs (see U.S. Census Bureau, 2002: Chapter 8; also available at <http://www.census.gov/prod/cen2000/doc/sf3.pdf>).

The Census Bureau is producing MOEs for every ACS estimate. The Census Bureau's current strategy is to make it clear to data users that there is statistical sampling error associated with estimates produced from samples. It was a lot simpler back in the SF3 days to just take the census long-form numbers as the "truth" and not bother with the details of sampling error and levels of uncertainty. However, what the Census Bureau is doing now with the ACS estimates is actually better and more informative. If you plan to compare Census 2000 long-form estimates with your recent ACS estimates to see if they are statistically different, there are many caveats to consider, and I'll continue to bring these up rather often, I'm afraid.

Computing the MOE for a Census 2000 sample estimate is not difficult.

[November 18] Clerk Smith: Do indeed bring them up whenever you think it's helpful. But can you give me an example of how I can compute an MOE on a Census 2000 SF3 estimate?

[November 21] Demographer Adams: Okay, I worked on this a bit over the weekend, and to help get you started I have computed the 90 percent MOE for the Census 2000 SF3 long-form sample estimate of the number of Wolf Lake's civilian workers engaged in private wage and salary (W&S) work. The data come from Table P51 in the Census 2000 SF3 for Wolf Lake. This is the "Class of Worker" table. Private W&S workers include all civilian employees (age 16 and over) who don't work for the government, aren't self-employed, and aren't unpaid family workers. If you look at Table P51 in the American FactFinder, you'll see that it's a very large table with a lot of detail. I combined male and female workers and summed across several cells in the table to get a single total for W&S workers.

To calculate the Census 2000 MOEs, you need four things:

- The unadjusted standard error or SE for the characteristic that would result from a simple random sample.
- The design factor for this characteristic to account for the Census 2000 long-form sample design (which wasn't a simple random sample).
- The Census 2000 sampling rate for the geographic area.
- The total population of the area.

Refer to "Example 1" in Chapter 8 of the Technical Documentation for SF3 (U.S. Census Bureau, 2002) to calculate the unadjusted SE for a total. Calculating standard errors for other kinds of estimates (e.g., percentages, means, and medians) are also shown in Chapter 8. The formula for the adjusted SE for a total is

$$SE_{Census\,2000} = \sqrt{5\hat{Y}\left(1-\frac{\hat{Y}}{N}\right)}$$

Where $\hat{Y}$ = the estimate of the characteristic total and
N = the size of the publication area.

The value of 5 in this formula assumes a sampling rate of 1-in-6. We'll have to adjust for this later since Wolf Lake was sampled at a higher rate. I found the observed Census 2000 sampling rate for Wolf Lake in SF3 Table P4. It was 44.17 percent.

In our example, the estimate of the characteristic total is the estimate of the civilian population 16 and over who are private wage and salary workers. From Table P51 in SF3, I found that estimate to be 335. The size of the publication area is the total population of Wolf Lake, 958. So the unadjusted SE is 33.00.

Refer to Appendix 3 for more information about margins of error and confidence levels.

$$SE_{Census\,2000(unadjusted)} = \sqrt{5(335)\left(1-\frac{335}{958}\right)} = 33.00$$

The design factor adjustment of 0.5 can be found in Table C in this technical documentation. Note that these design factor tables are state specific. I found "Class of Worker" in the stub of our state's table and then chose the final column "35 percent or more," given that the Census 2000 long-form sampling rate for Wolf Lake was 44.17 percent. Multiplying the unadjusted SE by the standard error design factor from Table C gives us an adjusted SE of 16.50 persons.

$$SE_{Census2000(adjusted)} = 0.5(33.00) = 16.50$$

I calculated the 90 percent margin of error for the estimate by simply taking the adjusted SE and multiplying it by 1.645, the factor for a 90-percent confidence level.

$$\text{MOE}_{(90\%)} = 1.645(16.50) = 27.14$$

This 90 percent MOE is used to construct a 90-percent "confidence interval" around the Census 2000 long-form estimate—308 to 362, and this now can be compared to the 90-percent confidence intervals associated with the ACS 5-year estimate. I found the 90 percent MOEs for the ACS estimates on the Census Bureau's Web site and created the corresponding confidence intervals. I found that interval to be 526 to 620.

When comparing the difference between two sample-based estimates, formal tests for statistical significance of the difference should be made. Appendix 4 presents a step-by-step procedure for making such comparisons.

[November 22] Clerk Smith: Since those intervals don't overlap, can I conclude with certainty that the ACS estimate is higher than the Census 2000 sample estimate?

[November 22] Demographer Adams: That's a tempting conclusion to draw, but you shouldn't do so before dealing with matters of "statistical significance." The number of W&S workers seems higher in the ACS estimate, and this makes sense given what surely must be the higher number of W&S workers in the summer months when more of the lodging establishments, restaurants, bars, bait & tackle shops, and so forth are open. I calculated a formal test of significance on the difference between the census estimate (335) and the ACS 5-year estimate (573), and the difference is statistically significant at the 90-percent confidence level.

The number of completed interviews used to create the ACS 5-year period estimates will generally be smaller than the number used to produce the Census 2000 long-form sample estimates.

Note also that the 90-percent confidence interval for the ACS estimate is wider than the one I calculated for the Census 2000 long-form estimate. This is to be expected given that the 5-year accumulated ACS samples for Wolf Lake are somewhat smaller than the Census 2000 long-form sample—something that's true for virtually all census geography.

[November 23] Clerk Smith: This whole discussion of sampling errors has been very helpful. Thanks, Terry. I've looked at Chapter 8 of the SF3 Technical Documentation and generated standard errors for everything in my report that came from SF3. I revised my original table to reflect those estimates, and it's clearer to me now that some of the estimates that looked very different were really only artifacts of sampling error. I can see now the value of looking at sampling errors before drawing conclusions. Can you look at this table and tell me what you think? Also, on November 18 you said in your message that you wanted to talk about a couple of issues related to sampling errors in Census 2000 and sampling errors in the ACS. Did I miss the second point?

Table A (revised). **A Few Wolf Lake Characteristics**

	Census 2000	2005–2009 ACS	2010 Census
Total housing units	1,407	1,587 (±60)	1,474
Occupied housing units	445	624 (±33)	452
Vacant housing units	962	963 (±53)	1,022
Percent vacant	68.4	60.7 (±0.3)	69.3
Owner-occupied housing units	358	405 (±28)	361
Total population	958	1,545	1,006
Total civilian private wage and salary workers	335 (±27)	573 (±48)	NA
Employed in retail trade	70 (±15)	127 (±26)	NA
Employed in arts, entertainment, recreation, accommodation, and food service	89 (±17)	399 (±29)	NA
Median age	49.6	39.5 (±2.8)	49.2
Percent age 65 and over	25.7	22.3 (±2.1)	24.3

NA Not Available.

ACS estimates for rural areas and small communities are based on proportionately larger samples than in large places.

[November 24] Demographer Adams: Your table looks great. I'm glad you are feeling more comfortable using MOEs and no, you didn't miss the "other issues." In my long-winded response, I did. I apologize; there's so much to say on this topic! The second point I wanted to make about sampling errors actually takes us back to my comments on November 10 regarding differences between the census long-form sampling approach and that of the ACS. While there are differences between the ACS sample and the Census 2000 long-form sample, there's actually one commonality that you'll appreciate. For small governmental units like Wolf Lake, the census long-form questionnaire was sent to a higher proportion of housing units than in larger cities. The Census Bureau did this so that the data obtained from the long-form sample for small areas would achieve a level of statistical reliability comparable to that of larger cities. The ACS survey follows a similar pattern. It, too, oversamples addresses in small communities, but, as we've discussed, not at a rate high enough to produce statistically reliable 1-year or 3-year estimates.

Despite the increased sample size for small governmental units, the ACS currently is producing single-year results only for governmental areas and statistical areas with populations at least 65,000. For counties and other geographic entities with populations ranging from 20,000 to 65,000, ACS estimates require the pooled results of 3 years' worth of ACS data. For governmental units like Wolf Lake, and for county subdivisions with fewer than 20,000 persons, ACS estimates require the pooled results of 5 years' worth of ACS samples. These ACS 5-year estimates are what you're looking at, and it takes some time to understand what these are, what they mean, and how they can be interpreted and used.

Assigning an ACS multiyear period estimate to the midpoint of the collection period is appealing but not a good idea.

For example, on November 7 you asked me what the 5-year estimates mean. I interpreted your question to be whether these period estimates can somehow be assigned to a specific year. The time frame described by multiyear estimates is the set of years included in the full range or period. The 2006–2010 ACS estimates are based on data collected in 2006, 2007, 2008, 2009, and 2010 and they therefore describe the characteristics of population and housing across those years. The ACS multiyear period estimates are not a simple average of several annual estimates. Some people think of these estimates as approximating conditions in 2008, recognizing full well that this midperiod estimate may not reflect at all the actual conditions in 2008. But this isn't everyone's preference—the Census Bureau encourages data users to think of the estimates as covering the full period of time.

[November 25] Clerk Smith: Thanks. That really does help me understand these numbers. As you know, we had an employer shut down his small manufacturing plant in May of 2008. Forty-two people lost their jobs overnight. We're still recovering from that event. How will the ACS results measure this change in the employment data?

ACS multiyear estimates will not be able to identify abrupt changes in characteristics. These estimates will smooth the data gathered prior to the event with data gathered after the event. Consult Appendix 1 for additional detail.

[November 28] Demographer Adams: One of the good things about the ACS is that it is designed to respond to such changes almost immediately, so that these kinds of abrupt events in large cities will show the effects of the event from one year's estimates to the next. The ACS samples prior to May 2008 probably picked up employees of that plant. Samples after May, of course, did not. Regrettably, for small places like Wolf Lake, the 5-year averages will smooth out what was, in fact, an abrupt moment for your community. Again, your ACS estimates are based on data collected over a 5-year period. I know that you wish that this abrupt change in manufacturing employment could be clearly reflected in your data, but it cannot. You certainly can, however, report the ACS data in your comprehensive plan and boldly footnote all employment data to indicate that the manufacturing plant closing affects them.

[November 30] Clerk Smith: And what about the annexation I told you about? We acquired several acres of land from Mason Town in November 2006. The county surveyor said we picked up 44 additional housing units in the process.

All ACS estimates use geographic boundaries existing on January 1 of the final year of the survey collection period.

[December 1] Demographer Adams: I'm glad we came back to that, since this is an important aspect of ACS data. The Census Bureau regularly updates geographic boundaries of governmental units using its annual Boundary and Annexation Survey. No doubt you've had personal experience in responding to that survey over the years. As you probably also know, the Census Bureau sends out a request for school district boundaries to be updated every 2 years. Boundaries of statistical areas such as census tracts are updated each decade as part of the decennial census process.

If the boundaries of a governmental unit change over the ACS estimation period (e.g., the result of an annexation like yours), the Census Bureau prepares the multiyear estimates using the geographic boundaries as of January 1 of the final year in the estimation period. The 2005–2009 estimates reported for Wolf Lake include the annexation because they're reported for the town using boundaries as of January 1, 2009. The data from the earlier survey years have been adjusted for the annexation that occurred even after those surveys were out of the field and closed. In other words, if some of the 44 housing units had fallen into the ACS sample in February of 2006, the characteristics of those housing units and the people staying in them would initially have been assigned to census geography associated with Mason Town. By the time the 2005–2009 ACS estimates were released, however, those housing and person characteristics would have been reassigned to Wolf Lake's data set.

[December 2] Clerk Smith: Okay, that seems clear enough. Now, I hate to ask this next question on a Friday afternoon, but my sense was the other day (the 24th) that you still wanted to say more about those margins of error. There's more?

ACS period estimates are affected by both sampling and nonsampling error. The appendixes attached to this handbook contain additional information about these errors.

[December 5] Demographer Adams: There's always more! Remember I teach this material and what I said on November 24 really wasn't enough. I think what you're asking is, "How good are these estimates?" So let me pretend I'm talking to the students in my demography class for a moment. For them, I would address the issue in two parts: sampling error and nonsampling error.

Sampling error derives from the fact that the data are based on a sample of the population rather than the entire population.

First, sampling error. The nature of inaccuracies that are associated with estimates based only on a sample of all possible observations is well understood and lies at the core of classical statistical theory. These errors are stated in terms of probabilities. When a sample of a population is taken, and an estimate of some attribute of that population is made, such as an estimate of the percent of all households having exactly 2 persons, the estimate will differ from the true unknown percent. The difference that can be attributed to selecting a sample is called "sampling error" or sometimes "sampling variability." Knowing exactly how the sample was drawn, or the "sample design," and how many households were sampled, allow us to establish a range above and below the estimate, known as a "confidence interval," that is likely at a given level of confidence to contain the true percent. Said another way, statistical theory can tell us how often the confidence interval will include the true unknown percentage. For all estimates derived from the ACS, the Census Bureau reports a margin of error at a confidence level of 90 percent. So, when you see an estimate in an ACS table of 40 percent ± 10 percent, you know that statistical theory has been used to guide your understanding of the estimate, i.e., your level of "trust" in the estimate. It tells you that the interval 30 percent to 50 percent has a 90 percent probability of containing the true percentage.

The strategy of the Census Bureau to cumulate many months of data before providing estimates based on the ACS for small communities like yours is driven by a goal to provide estimates with reasonable confidence intervals. In formal statistical terms that also have intuitive meaning, estimates with narrower confidence intervals are called "more reliable" or "more precise."

Nonsampling error includes a broad set of errors resulting from data collection or data processing problems.

Now, let's discuss nonsampling error. In addition to sampling error, for which rote formulas are available to compute it, there are other reasons why an estimate may differ from the true, but unknown, value. Nonsampling error can occur for a number of reasons and the consequences of this type of error are often poorly understood. Such error can result from missing data, response errors, or inadvertent errors in data collection or data processing. Some of these errors arise from Census Bureau procedures; others arise in the household as respondents deal with the ACS questionnaire. The ACS provides measures on its Quality Measures Web site <http://www.census.gov/acs/www/UseData/sse/index.htm> and detailed quality measures tables as data products in American FactFinder that indicate the potential for some types of nonsampling error.

Nonsampling error in the ACS is smaller than nonsampling error associated with the Census 2000 long-form survey.

Okay, one more important point to make about nonsampling errors. Permanent, highly trained Census Bureau employees carry out the fieldwork for the ACS. Temporary field workers, often with minimal training given the speed with which the census fieldwork must be accomplished, did the decennial census fieldwork, including gathering much of the long-form data. As a consequence, ACS nonsampling error is lower than nonsampling error affecting the long-form sample data. That's very good news!

Enough. But do check out the Quality Measures Web site when you get a chance. This is part of the process when using ACS data. Class adjourned.

The Census Bureau provides information about the sample sizes used to produce ACS estimates.

[December 6] Clerk Smith: Actually, that helped a lot. But here's another issue. A town resident called me yesterday to say that she had received a letter telling her that her household was going to receive a questionnaire in the mail for the American Community Survey. Boy was I ready for this call! We talked about the ACS for a while, and she promised me she'd fill it out for herself and her husband. Here's my question. How many people in Wolf Lake are likely to be interviewed for the ACS over the course of a year?

An address can only be selected for the ACS once every 5 years.

[December 6] Demographer Adams: One of the detailed tables produced from the ACS and available through American FactFinder (B00001) summarizes the unweighted sample count of the population for every geographic area for which data are published. This is the count of the total number of people in the sample that were used to produce the ACS estimates. For Wolf Lake the 5-year total of 624 tells you the total sample over 5 years. If we divide that by 5, we get an approximation of 125 people in sample each year in Wolf Lake.

Oh, you might also tell the woman asking about her address being selected this month for ACS interviewing that the Census Bureau will ensure that her address isn't sampled more than once every 5 years.

[December 7] Clerk Smith: You didn't mention the nursing home. What about the people who live there?

Persons living in group quarters are also included in the ACS sample (since 2006).

[December 7] Demographer Adams: That estimate we just came up with is for the total population and would include any people who were interviewed in group quarters facilities (e.g., college dormitories, nursing homes, and prisons). Residents of your nursing home may not be interviewed every year but over a 5-year period, the ACS estimates will probably include data for people living in this facility.

[December 9] Clerk Smith: Terry, I'm beginning to put together some tables for our planning exercise, and I meant to mention this much earlier. I really don't understand some of what I am seeing in the ACS data. Our Census 2000 population was 958 and the 2010 Census count released with the redistricting file this past March pegged our population at 1,006. The demographic office in the state government estimated our population at 1,002 in 2007—entirely consistent with the census counts. The 2005–2009 ACS estimate of 1,545 just doesn't fit with the others. What's going on here?

Table B. **Population Counts and Population Estimates for Wolf Lake**

	April 1, 2000	July 1, 2007	2005–2009 ACS	April 1, 2010
Census Count	958			1,006
State Demographic Estimate		1,002		
Census Bureau Estimate		993	1,545	

Because independent population estimates are used to produce ACS estimates of total population, ACS estimates will usually be consistent with other census-based estimates at the county level.

Population estimates for subcounty geography reported in the ACS may differ from other population estimates prepared by the Census Bureau or by state demographic offices.

[December 12] Demographer Adams: We touched on this issue last month. What you are seeing here is a difference that some small towns may see between the ACS period estimates, the census counts, and other population estimates.

In most cases, the Census Bureau forces ACS 1-year estimates of total population at the county level to match the estimates made by the Census Bureau's Population Estimates Program. If you were to look at the 2007 ACS 1-year estimates of total population for Lake County, they probably would be reasonably close to the 2007 Population Estimates Program estimates. Most of the time, and for most counties, these procedures bring everything into alignment. While those independently derived county estimates aren't perfect by any means, the numbers probably look like you would expect them to look.

There are situations, however, when these procedures produce ACS period estimates for small towns and villages that may look very different from estimates from both the state and the federal estimates programs. I think you've spotted one such instance, as this can happen with ACS estimates of population for subcounty units of government. For subcounty units like Wolf Lake, even the estimate of total population is derived as a *sample-based* estimate. There are lots of reasons why these subcounty estimates from the ACS will differ from other available subcounty estimates. If, for example, a geographic entity is affected by seasonal population swings, as Wolf Lake certainly is, the ACS estimate of total population can differ considerably from the most recent census count or the independent population estimate made by the Census Bureau or by a state statistical agency. This could happen in small college towns or resort and tourist communities like Wolf Lake. Small rural communities that bring in migrant farm workers for seasonal harvesting may also see differences from census estimates. The Census Bureau is aware of this and the problems that some users, like you, could have. They are currently researching options that would involve the use of subcounty population estimates in the ACS weighting in hopes that this would minimize these types of differences. But, in the short term, you should be prepared for these differences to exist. Stay tuned!

It's also important to recognize that the ACS 5-year estimates are not forced to match any single set of independent estimates. Instead, the Census Bureau uses a simple average of the independent estimates from each of the 5 years of the period. For this reason, you shouldn't expect the estimate of total population in the 5-year estimates to match estimates from any given year's Population Estimates Program exactly, even at the county level.

I'll make one additional comment about this. The ACS was designed to provide characteristics of the population, not estimates of total population or housing units. While they are released with the ACS data, I recommend that, when available, you use the state or Census Bureau independent population estimates whenever you need this kind of information. You will have to use the ACS data for an update of the other characteristics for your report, and to do this you should express both the census and ACS estimates as percentages. You will probably see differences between the Census 2000 and the ACS percentages, and you should do the requisite statistical testing of these differences as we have described to determine if any of the differences are just the result of sampling error. You will have to bring together everything we have talked about to do this, but give it a try. Good luck. Be in touch when you have your tables and we can review them together.

[December 13] Clerk Smith: I agree that I should probably use our state demographic agency's intercensal population estimates, but I anticipate the ACS estimates of population could cause us some trouble. We receive a sizable payment each year through the state's revenue sharing program. This payment, in excess of $50,000 for Wolf Lake in recent years, is largely based on our population size. Population size also determines our eligibility for other state and federal programs. And as fate would have it, a population size of 2,000 is the threshold level for some of these programs. The ACS estimates are inching toward the 2,000 mark. Our town board has already asked me if we should challenge the state demographic office's estimates (and the Census Bureau's own estimates!) based on the higher ACS estimates. When the ACS places us over the 2,000 level, which it appears could happen in a few years, there's going to be a lot of questions regarding which estimate is the correct one!

ACS estimates of total population cannot be used to challenge other population estimates—for example, those estimates produced by the Census Bureau's Population Estimation Branch or other state agency population estimates.

[December 13] Demographer Adams: There's not much more I can say about this, Pat. I have been told that ACS estimates of total population cannot be used in population estimates challenges. But I agree that there may be some important questions later on. You at least have the advantage of now understanding why this is happening.

[December 14] Clerk Smith: I'm not going to worry about that now. At the moment I'm getting some serious heat from the county to finish my work before the end of the year. They're asking me for my draft tables ASAP, and my own town board is also pressuring me to get this off my desk and move on to other things.

Some of my draft tables are shown here. I'm working on a lot of other tables, but I thought I would share these with you to get some feedback. I've broken the topics into several small tables because, for our final report, that's the way the data will be shown: small tables with associated commentary. As you'll see, the ACS data feature rather prominently in my tables. As usual, your comments will be gratefully appreciated.

Table 1. **Total Population—Town of Wolf Lake**

Location	Census 2000	2007 Intercensal Estimate	2008 Intercensal Estimate	2010 Census
Town of Wolf Lake	958	993	993	1,006
Share of Lake County	4.6%	4.5%	4.5%	4.5%
Lake County	21,033	22,047	21,919	22,361

Table 2. **Population by Age—Town of Wolf Lake**

	Census 2000	2005–2009 ACS	2010 Census
Total population	958	1,545	1,006
Percent	100.0	100.0	100.0
Less than 5	3.8	2.7 (±0.7)	3.1
5–9	4.9	3.4 (±0.8)	4.4
10–14	5.1	3.9 (±0.8)	5.3
15–19	5.1	12.2 (±1.3)	6.2
20–24	2.3	10.7 (±1.8)	3.9
25–34	7.3	10.4 (±1.6)	7.1
35–44	12.9	13.7 (±1.4)	12.9
45–54	16.5	19.2 (±1.7)	17.1
55–59	7.7	2.9 (±0.4)	7.3
60–64	8.7	1.7 (±0.4)	8.2
65–74	14.8	7.8 (±1.1)	14.0
75–84	8.2	8.9 (±1.8)	8.1
85 years and over	2.6	2.6 (±0.6)	2.4

Table 3. **Population by Race—Town of Wolf Lake**

	Census 2000	2005–2009 ACS	2010 Census
Total population	958	1,545	1,006
Percent	100.0	100.0	100.0
One race	99.3	98.1 (±0.7)	98.6
White	98.2	98.1 (±0.7)	98.1
Black/African American	0.0	0.0 (±1.6)	0.0
American Indian or Alaska Native	0.5	0.0 (±1.6)	0.3
Asian	0.1	0.0 (±1.6)	0.0
Native Hawaiian or other Pacific Islander	0.0	0.0 (±1.6)	0.0
Some other race	0.4	0.1 (±0.1)	0.2
Two or more races	0.7	1.8 (±0.7)	1.4
White and Black	NA	0.0 (±1.6)	0.0
White and American Indian or Alaska Native	NA	1.8 (±0.7)	1.4

Table 4. **Household Income—Town of Wolf Lake**

	Census 2000	2005–2009 ACS
Income less than $25,000	36.6 (±3.5)	19.9 (±6.1)
Income less than $25,000 to $99,999	59.2 (±3.8)	73.8 (±6.7)
Income greater than $100,000	4.2 (±2.0)	6.3 (±3.5)
Households with earnings	68.6 (±3.7)	75.5 (±6.5)
Households with social security income	44.3 (±3.7)	35.1 (±6.5)
Households with retirement income	18.9 (±2.9)	11.9 (±5.1)

Table 5. **Year Structure Built—Town of Wolf Lake**

	Census 2000	2005–2009 ACS
Total Housing Units	1,407	1,587
Percent	100.0	100.0
Built 1980 or later	35.2 (±1.4)	27.9 (±2.5)
Built 1960 to 1979	26.6 (±0.9)	28.7 (±1.6)
Built 1959 or earlier	38.2 (±1.5)	43.4 (±2.6)

[December 17] Demographer Adams: Pat, you've really done a wonderful job. This is a great beginning. In general, I like the way you've presented the ACS data showing the margin of error. I checked a few of the 90 percent MOEs you've calculated for numbers based on the Census 2000 long-form sample, and those look correct.

ACS estimates won't always be the best source.

Table 1. This is good. The best source of estimates of the total population should come from the decennial census, the Population Estimates Program, or other state-based projections. The ACS is not the best source for this information.

When ACS sample estimates have large MOEs, it's probably time to sacrifice detail and collapse the cells of the table in order to obtain estimates based on a larger sample size.

Table 2. Since you have the census counts by age both for 2000 and 2010, I think I would just use those data. You have the basic demographics from the 2010 Census with which to update the town's demographic data, so leaving the ACS data out of these tables would be fine. There's no right or wrong way. I'll leave it up to you. You know your community and the Lake County planners who will be analyzing your report better than I. If you leave in the ACS data, you may want to consider collapsing some of the smallest age categories to improve the reliability of those estimates. You will also need to explain why some of the age distributions are so different.

Table 3. My comment here is similar to that for Table 2. The mingling of census and ACS data sometimes isn't necessary. The ACS is designed to give small communities detailed estimates of characteristics that are not picked up in the census. Since you have 2000 and 2010 Census results for race, there's no need to also show the ACS 5-year estimates.

Use ACS estimates to show the percent distribution of characteristics of your community's population.

Table 4. This is great. It shows exactly what the ACS is for—to give you updated characteristics (in percentage terms) of items not covered in the census. You know how to compute statistical significance on differences based on our discussion on November 22. The higher incomes in the ACS may be telling you that the summer residents have higher incomes than the residents that are included in the census. The ACS data could also be indicating income increases in Wolf Lake since 2000.

Table 5. Again, I like what you did. I see you collapsed some of the detailed categories—that's a good idea. In general, people have trouble recalling the specific year that older homes were built so combining homes built 1939 or earlier with those built 1940 to 1959 is probably giving you better estimates anyway. The ACS data provide you with refreshed information on the housing stock of Wolf Lake.

[December 17] Clerk Smith: I'm glad to hear you say that they look reasonable. I couldn't have ever done this without you, Terry. You've helped me develop confidence in using these new data and I hope I can help others do the same.

The ACS will reveal information about your community that will no longer be available from the census.

[December 17] Demographer Adams: I have just a few general final comments. First, there's no question that the ACS is adding richness to the community profile of Wolf Lake. It's documenting the characteristics of your community on a regular basis. Without a long form in the 2010 Census, you need to get comfortable with the ACS.

The ACS will permit a community to compare itself with its neighbors every year.

Second, as I looked at these estimates, I would love to have compared them with some of the other small towns in Lake County or with the corresponding tables for Lake County itself. Such comparisons really add context to the data tables you've created for Wolf Lake. I do understand that the other towns are also preparing reports like yours and that the county planners are preparing a report for the county. Since these all follow a basic tabular template, the comparisons will be easy to make. But, again, the beauty of the ACS is that you can track changes over time with respect to how Wolf Lake is changing relative to your neighbors and relative to the county as a whole.

I'll say it once again—you've done a terrific job! Good luck with the next stage of this project!

Epilogue

This 5-week conversation between Clerk Pat Smith and Demographer Terry Adams is entirely fictional. As a hypothetical small community invented for purposes of this handbook, Wolf Lake doesn't really exist, but the kinds of issues highlighted in this handbook are quite real. Not all small communities in the United States are like Wolf Lake with its ACS estimates simultaneously influenced by seasonal swings in population, the abrupt closing of a major local firm, and geographic boundary changes. The combination of those forces was included in order to talk about them. But most small rural communities and small cities likely share at least one of these attributes, and for some such places, the attribute could be more extreme than illustrated here for Wolf Lake.

It's not clear how many communities share with Wolf Lake the influx of seasonal populations for 3 to 4 months of the year. These communities include places, like Wolf Lake, where the seasonal peak occurs in the summer as well as places that have their seasonal peak in the winter months. There are many small towns with colleges and other institutions of higher learning where a common pattern of seasonality is shared. The same can be said about agricultural communities who rely on migrant farm workers to harvest and process seasonal agricultural commodities.

In most aspects of our hypothetical dialogue presented here, Clerk Smith makes effective and good use of the ACS data. Her approach and her experience level are likely typical of small area data users interested in the characteristics of their communities. She dug into the technical documentation to learn about the ACS, and, wisely, she made excellent use of resource people to assist her in working through the ACS data and their complexities. Virtually all small community data users have access to demographic experts in their own state's State Data Center agency at <http://www.census.gov/sdc/www/>. These agencies have experts on staff like Demographer Adams and are happy to work with data users like Clerk Smith.

Reference

Citro, Constance F., and Graham Kalton (eds.). 2007. *Using the American Community Survey: Benefits and Challenges.* Washington DC: The National Academies Press.

U.S. Census Bureau. 2002. *Census 2000 Summary File 3 Technical Documentation.* Washington DC: U.S. Census Bureau.

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 U.S., 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, United States 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.

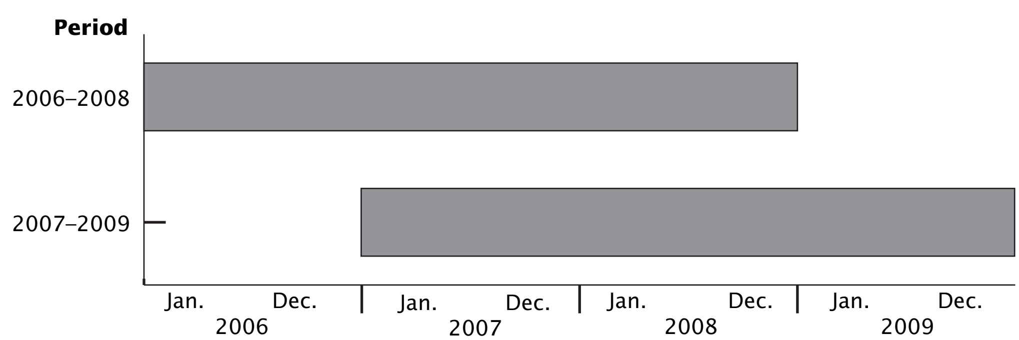
You certainly can compare characteristics for areas with populations of 30,000 to areas with populations of 100,000 but you should use the data set that they have in common. In this example you could use the 3-year or the 5-year estimates because they are available for areas of 30,000 and areas of 100,000.

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**

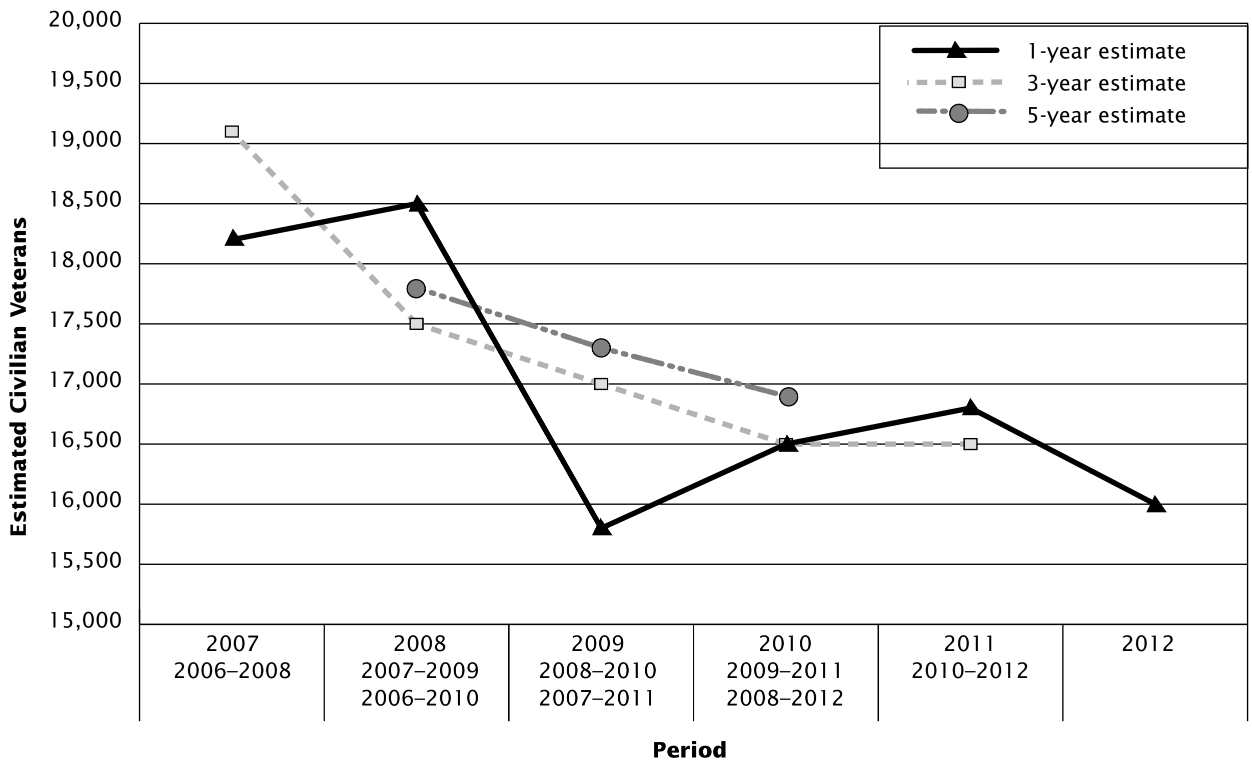


Source: U.S. Census Bureau.

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.

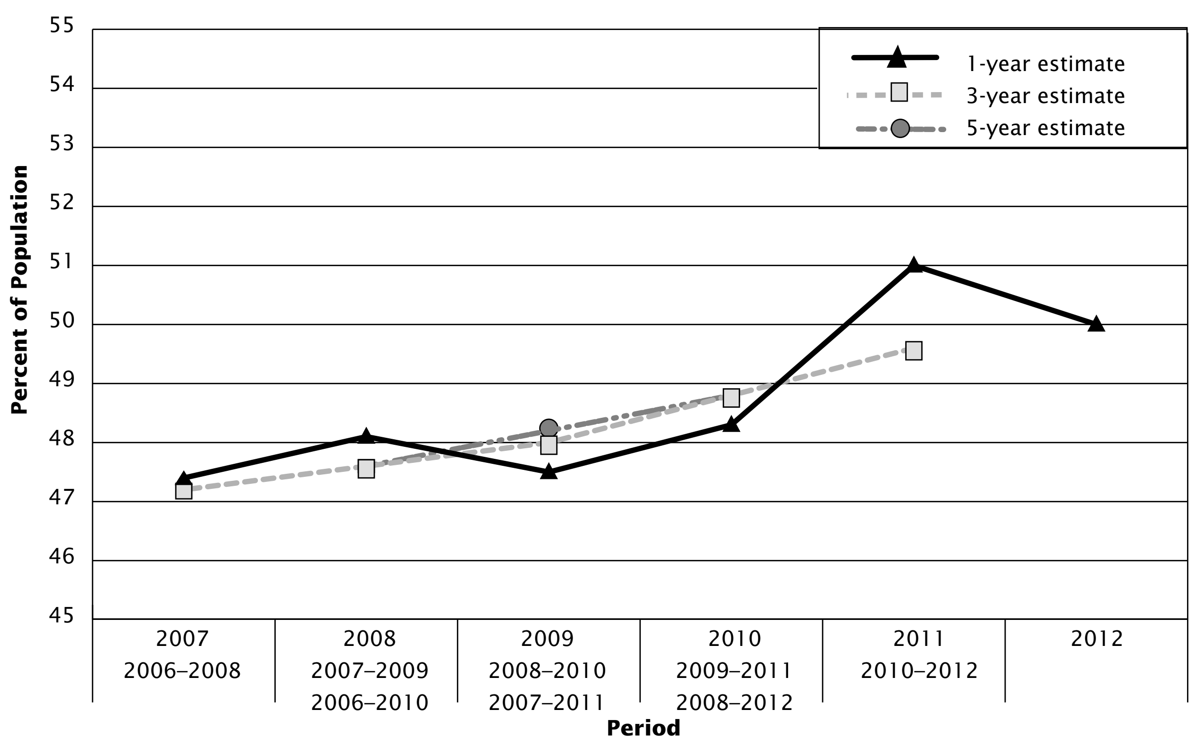
Local area variations may occur due to rapidly occurring changes. As discussed previously, multiyear estimates will tend to be insensitive to such changes when they first occur. Single-year estimates, if associated with sufficiently small CVs, can be very valuable in identifying and studying such phenomena.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

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

Deriving the Standard Error From the MOE

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

Derivation of SEs can thus be expressed as

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

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

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

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

Confidence Intervals

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

Constructing a Confidence Interval From a Margin of Error

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

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

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

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

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

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

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

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

The CI can thus be expressed as the range

$$CI_{CL} = (L_{CL}, U_{CL}).$$ [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).

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

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

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

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).

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

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

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

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

Standard Error (SE)

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

Confidence Interval (CI)

$$CI_{CL} \quad \left(\hat{X} -= MOE_{CL}, \hat{X} += MOE_{CL}\right)$$

Coefficient of Variation (CV)

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

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:

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

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

MOE_{den} is the MOE of the denominator.

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

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

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

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

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

Table 2. **Data for Example 2**

Characteristic	Estimate	MOE
Black females 25 years and older with a graduate degree (numerator)	4,634	±989
Black females 25 years and older (denominator)	31,713	±601

The estimated proportion is:

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

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).

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

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

$$MOE_{num}^2 - (\hat{p}^2 * MOE_{den}^2)$$

$$(989)^2 - [(0.1461)^2 * (601)^2]$$

$$978,121 - 7,712.3 \quad 970,408.7$$

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

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

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:

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

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

MOE_{den} is the MOE of the denominator.

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

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

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

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

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

The estimated ratio is:

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

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).

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

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

$$MOE_{num}^2 + (\hat{R}^2 * MOE_{den}^2)$$

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

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

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

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

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:

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

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

MOE_A is the MOE of the first estimate.

MOE_B is the MOE of the second estimate.

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

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

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

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

$$\pm 1{,}405$$

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

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

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).

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:

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

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

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.

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

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:

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

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:

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

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:

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

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:

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

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.