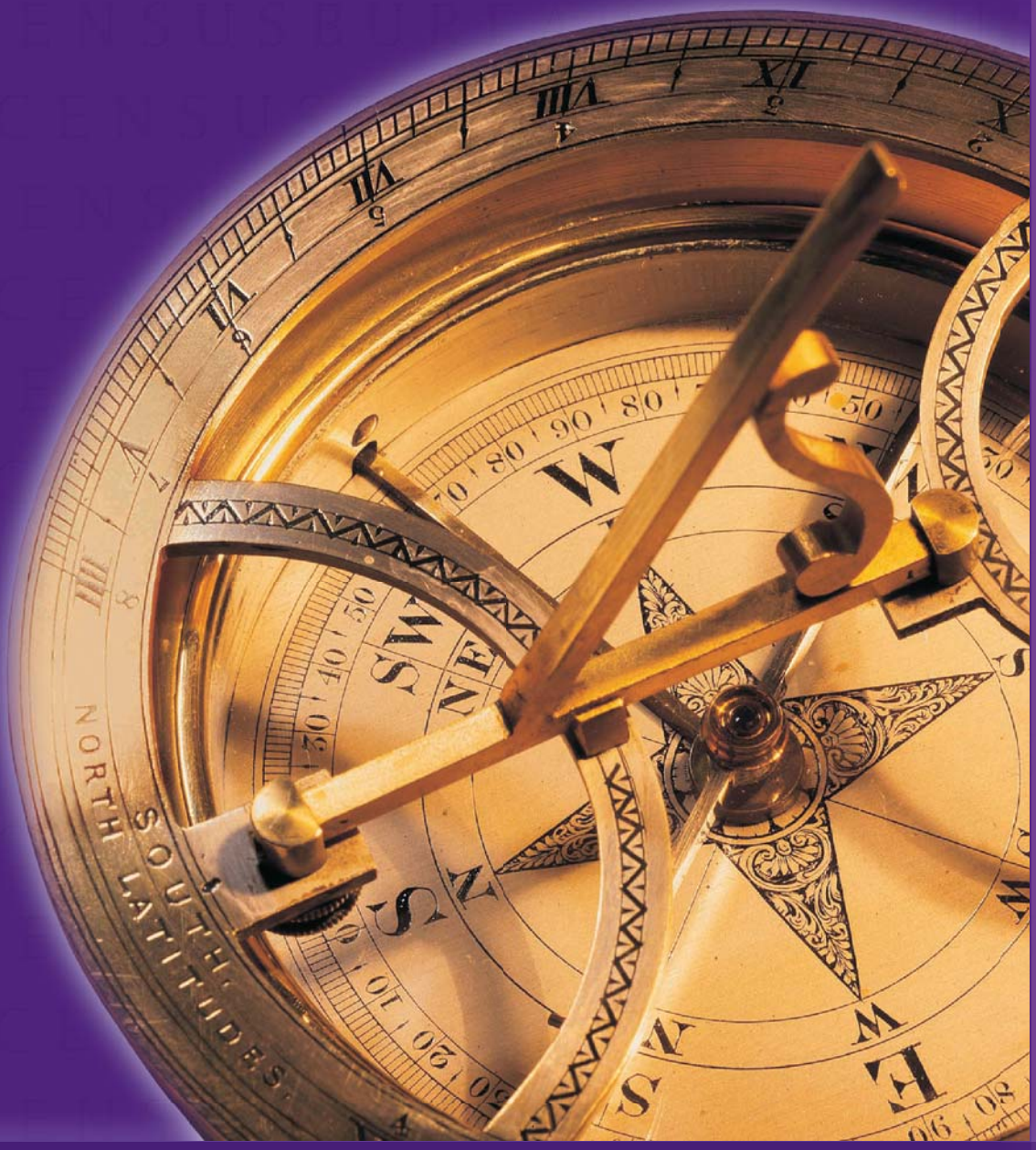
A Compass for Understanding and Using American Community Survey Data

What State and Local Governments Need to Know

Issued
February 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

Joseph J. Salvo, **Arun Peter Lobo**, and **Joel A. Alvarez**, Population Division, New York City Department of City Planning, 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.

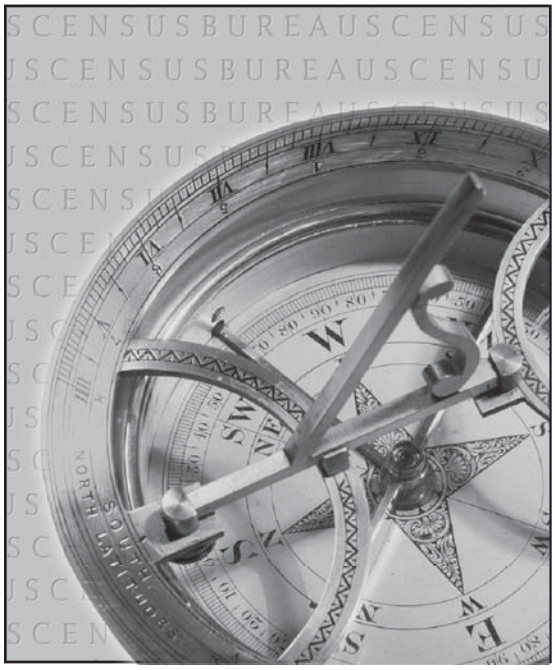
Other individuals who contributed to the review and release of these handbooks include **Dee Alexander**, **Herman Alvarado**, **Mark Asiala**, **Frank Ambrose**, **Maryam Asi**, **Arthur Bakis**, **Genora Barber**, **Michael Beaghen**, **Judy Belton**, **Lisa Blumerman**, **Scott Boggess**, **Ellen 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 February 2009

What State and Local Governments Need to Know



U.S. Department of Commerce
Vacant,
Secretary

Vacant,
Deputy Secretary

Economics and Statistics Administration
Vacant,
Under Secretary for Economic Affairs

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

Suggested Citation

U.S. Census Bureau,
A Compass for Understanding and Using American Community Survey Data: What State and Local Governments Need to Know
U.S. Government Printing Office, Washington, DC, 2009.

ECONOMICS AND STATISTICS ADMINISTRATION

Economics and Statistics Administration

Vacant,
Under Secretary for Economic Affairs

U.S. CENSUS BUREAU

Thomas L. Mesenbourg,
Acting Director

Thomas L. Mesenbourg,
Deputy Director and
Chief Operating Officer

Arnold A. Jackson
Associate Director for Decennial Census

Daniel H. Weinberg
Assistant Director for ACS and Decennial Census

Susan Schechter
Chief, American Community Survey Office

Contents

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

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

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

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

Introduction

A primary mission of state and local governments is to deliver efficient services and enact policies that advance public safety and economic growth. For decades, data from the decennial census long form have provided invaluable information that helped frame these issues in the proper context, enabling governments to create proposals, develop budgets, and execute strategies to address well-documented needs. With the elimination of the long form, state and local governments must avail themselves of the U.S. Census Bureau's American Community Survey (ACS), which will now provide detailed information about the population and housing attributes of states, counties, and municipalities, large and small. In addition, the ACS will offer state and local governments a more dynamic picture of their communities.

Unlike the decennial censuses, the ACS is a continuous national survey. Rather than collect data as of a single decennial reference point, the ACS collects data nearly every day and summarizes it over 1-, 3-, and 5-year periods. Much like the census, the ACS provides a picture of the social, economic, and housing characteristics of the population, but it has the added advantage of providing these data on a yearly basis. The provision of more current data is not without its tradeoffs, which include smaller samples and therefore higher levels of variability. In fact, ACS data users will be faced with compromises between the provision of current data and the reliability of estimates. The most significant of these choices will involve the use of estimates for small areas involving multiple years of data.

While differences between the decennial census and the ACS may be subtle, they frequently necessitate a new analytical approach. **The major goal of this handbook is to illustrate how ACS data can be used to address typical issues faced by state and local governments and, in the process, provide information that can facilitate an effective transition to ACS data**.

A glossary and a series of technical appendixes—for those interested in more advanced ACS applications—are included at the back of this handbook.

The Role of Data in the Delivery of Government Services

In the delivery of government services, data are often used to help establish priorities through a needs assessment, to develop general plans, and to implement selected plans. We briefly discuss how data are used in each of these steps. Often, this is not a linear process; rather these steps often feed into and inform each other.

Establishing Priorities Through a Needs Assessment

Given competing demands and limited resources at their disposal, governments need to carefully ascertain appropriate funding levels for their initiatives. ACS data can be analyzed to assess the level of need and to prioritize funding levels for proposed initiatives.

Governments also receive requests for help from myriad community groups and civic organizations that need to be assessed. ACS data could be extremely useful in evaluating the overall needs of the community and the size of local populations in need of various services in order to prioritize requests for assistance.

Developing a General Plan

Once a government decides on its priorities, it needs to come up with an effective plan. This often calls for examining various alternatives. If, for example, a local government decides to make the alleviation of poverty a priority, it needs to examine where exactly to apply its resources. Should the alleviation of child poverty be a priority or should the focus be on the elderly poor, or on the elderly poor who are living alone? Or, should resources be applied in some proportion to each of these groups? Examination of ACS data could be instrumental in formulating plans and actions to guide the distribution of resources.

Implementing the Selected Plan

Once a plan is decided on, it must be implemented. If, for example, a local government decides to focus primarily on the elderly poor, ACS data can be used to target neighborhoods for the delivery of funds and other resources. A specific program could be developed for neighborhoods with the largest number of elderly poor.

Important Points to Consider When Using the ACS

The ACS is based on a questionnaire that is sent each month to a sample of about 250,000 addresses in the United States.[1] Each calendar year, these data are pooled and estimates are produced for about 60 different social, economic, and housing characteristics. **Since the size of a geographic area largely determines the size of the sample, only larger areas—those with 65,000 or more people—receive 1-year estimates. For smaller places, estimates are created for multiyear periods: for areas with populations between 20,000 and 65,000, 3 years of data are needed; and for areas with fewer than 20,000 people, 5 years of data need to be collected in order to provide estimates**.[2] All of this differs from the decennial census, where the data are pegged to an April 1 reference point (even though data collection actually occurs into August of the census year).

The use of a continuous data collection operation means that the ACS can provide local governments with updates more than once a decade. Another innovation is that the ACS collects information for everyone who is living or staying in a sampled ACS unit for more than 2 months, even if they have another residence. (This is not the case with the decennial census, since "where you live most of the time" determines where you are enumerated.) For example, if a couple lives most of the year in Florida, but lives in New York City for 4 consecutive months each year, that couple may be included in the New York City ACS sample if their New York home is sampled and interviewed while they are residing there. Compared with the decennial census, the ACS will better account for seasonal migration patterns that occur in some communities and will measure the characteristics of the population actually present at various times of the year. For more information about how the ACS compares with the decennial census, see Appendix 2.

Estimates based on information collected over 1, 3, and 5 years are referred to as "period" estimates, meaning that numbers represent an area's characteristics for the specified period of time. Period estimates may be thought of as averages and represent the biggest conceptual change from the decennial census. Working with period estimates for 1, 3, and 5 years requires a different approach to data analysis, one that is best defined through the use of illustrations provided later in this handbook. Appendix 1 provides additional background on period estimates.

It is important to keep in mind that the main purpose of the ACS is to estimate the demographic, social, economic, and housing characteristics of the population. ACS estimates of the overall number of people and housing units at the county level are obtained from the Census Bureau's Population Estimates Program. Each year, the Census Bureau creates independent population estimates for all counties in the United States by evaluating change using a variety of primarily administrative data sources. Members of the Federal State Cooperative for Population Estimates (FSCPE) participate in the review and development of these estimates. For more information about the Census Bureau's Population Estimates Program, see their Web site at <http://www.census.gov/popest/estimates.php>. These detailed population estimates are used as survey controls to make ACS estimates of characteristics more reliable over time and to correct for deficits in the sample, i.e., differences in the level of coverage by age, sex, race, and Hispanic origin. Thus, while the ACS is the primary source of estimates on social, economic, and housing characteristics, overall population and housing estimates at the county level are independently determined by the Population Estimates Program and adopted by the ACS. Appendix 7 provides more information about the implications of population controls for ACS estimates.

Finally, like the decennial census long form, all ACS estimates for demographic, social, economic, and housing characteristics come from a sample. This means that a certain level of variability will be associated with ACS estimates. This variability is referred to as "sampling error" and is expressed as a band or "margin of error" (MOE) around the estimate. Understanding the basic methods of statistical sampling and the ramifications of working with sample data are key to using the ACS successfully. A detailed discussion of how to apply concepts such as MOEs is included later in this handbook and in Appendix 3.

Working With ACS Data

If ACS data are deemed appropriate, one has to decide on the right ACS product. Depending on the geographic area being examined, decisions need to be made as to whether a 1-year average would be more appropriate than a multiyear average. The availability

[1] About 3,000 addresses are also sampled each month for the Puerto Rico Community Survey. Similar methods are used to collect, process, and publish data for Puerto Rico.

[2] Each year a geographic area is evaluated to see if data for the area will be published in either or all of the 1-, 3-, and 5-year data products. An area is eligible for estimates as soon as it meets the total population threshold and for each year thereafter, as long as it does not fall 5 percent below the threshold. The release of estimates is always determined by the total population estimate for the latest year. An area that reaches 65,000 people in the latest year will receive a 1-, 3- and 5-year estimate, regardless of its population size in the prior years. For more information, please see Appendix 1.

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.

of these options requires data users to clearly define the objectives of their analyses, which is the single most important step to effective use of the ACS.

Table 1 summarizes, by type of geographic area, the proportion of the geographic area type that will receive ACS data in the form of 1-year, 3-year, and 5-year estimates; 3-year and 5-year estimates only; and those that will only receive 5-year estimates.

ACS Data Products

Data from the ACS are available in several forms, including data profiles, detailed tables, and Public Use Microdata Sample (PUMS) files, all accessible via the Census Bureau's Web site. The type of data needed is closely related to the question or issue being addressed. For many of the most basic applications, where single numbers (e.g., people below the poverty line) or summary statistics (median family income) are needed, data profiles are useful tools and are easily accessible. When data are needed for more detailed questions, such as the ratio of income to the poverty line or a distribution of income for households, detailed tables are frequently required. There are two levels of detailed tables, those that are abbreviated (these table numbers begin with a "C") and those with more detailed categories (table numbers beginning with a "B"). It is important to keep in mind that increased detail always comes at the expense of data reliability; the margins of error are generally larger for estimates in tables with the more detailed characteristics. Moreover, due to sample size constraints and the greater detail demanded, data may be too sparsely distributed in the cells of a detailed table, which may result in some tables being "blanked-out" or suppressed. Optimal data utilization occurs when there is

a good balance between the level of detail examined for a particular variable and the reliability of those estimates, as explained later in this handbook.

The most flexible data product is the PUMS. (See the text box titled "What Is the Public Use Microdata Sample?") Since microdata are actual ACS response records, these data can provide the most detailed tabulations, limited only by the number of records in the file. In exchange for this flexibility, the geographic codes provided for individual records are limited to states and specific areas of at least 100,000 population (called Public Use Microdata Areas or PUMAs). Aggregation of these records cannot occur for smaller geographic units, such as counties or census tracts.[3]

What Is the Public Use Microdata Sample?

Researchers who need tabulations that are more detailed than the Census Bureau's ready-made tables often use the American Community Survey's Public Use Microdata Sample (PUMS) file. Microdata files contain a sample of actual survey response records, minus any identifying information. This allows data users to create cross-tabulations using a host of variables. For example, a criminologist studying delinquency might want to know the number of young adults between the ages of 16 and 21 who are not enrolled in school, who have not graduated from high school, and who do not have a job. Exploring the American FactFinder, it is possible to get tables on this population only for those between the ages of 16 and 19. If the criminologist felt 20- and 21-year olds had higher rates of delinquency, he or she might want to use the PUMS to create a table for 16- to 21-year olds. There are tradeoffs, however. In order to preserve respondent confidentiality in these files, geographic identification is limited to states and a set of areas of 100,000 or more population, called Public Use Microdata Areas or PUMAs. Also, there are limits to the level of subject detail that PUMS can provide, based upon the number of people or housing units in the sample. More information on PUMS is available in the handbook for PUMS data users.

[3] PUMAs are combinations of contiguous census tracts. For more information about PUMAs and the Census Bureau's Public Use Microdata Sample (PUMS) files see the PUMS overview Web page at <http://www.census.gov/acs/www/Products/PUMS/index.htm> or the handbook in this series for PUMS data users.

Microdata products require more knowledge of the form and content of ACS data than other ACS data products and a clear idea of the purpose of the analysis; "browsing" table categories is not possible because there are no preset tabulations, as with the summary tables. Instead, the records need to be downloaded from the ACS Web site into software that can aggregate records using the variables in the file. Statistical software, such as SAS (Statistical Analysis System), SPSS (Statistical Package for the Social Sciences), or relational database software, such as MS ACCESS, can all be used to aggregate records to create tabulations or other output. Data users who do not have statistical software can elect to work with the PUMS files using an online data tool called DataFerrett. DataFerrett is a data analysis tool that provides access to PUMS files via the Internet. It can be installed as a stand-alone application or used through an Internet browser. The DataFerrett Web site also includes an online tutorial on how to access and use microdata files. While accessing PUMS requires additional work, data users can customize tables so that the results specifically address their needs.

Working With Multiyear Estimates

For data users who are working with places of at least 65,000 people, one by-product of the increased complexity of the ACS is the array of options that multiyear averages provide. A classic case in point is when a local government needs a number that represents the best "current" assessment of an attribute for its jurisdiction, or an estimate from the recent past. Let's assume the attribute of interest is public assistance.

There are multiple ways of measuring levels of public assistance using ACS data, ways that vary depending on the size of the geographic area being examined. Beginning in 2010, following the release of the first 5-year estimates, areas with 65,000 people or more will have the option of using 1-, 3-, or 5-year averages. The examples provided in the following sections are based on research from ACS test counties for which 1-, 3-, and 5-year data are currently available.

Choosing a Current Estimate

If one needs to examine the most current estimate of an attribute, one has to "anchor" the 1-, 3-, and 5-year estimates on the most recent year for which data are available. If the most current estimate of households receiving public assistance income is required, it would be ideal to use just the latest 1-year estimate. However, the 1-year estimate is derived from a smaller sample and is less reliable than 3- or 5-year estimates. So how does one decide on which estimate to use?

Along with estimates, the ACS provides MOEs, which allow users to calculate a standard error. When a standard error is divided by the estimate, the resulting statistic is known as the coefficient of variation or CV. (See the text box titled "What Are Margins of Error and Coefficients of Variation?") For example, if the number of people below the poverty line is 1,000 and the standard error is 80, the CV equals 8 percent, which indicates that the standard error equals 8 percent of the estimate. CVs are a standardized indicator of the reliability of an estimate. While there is no hard-and-fast rule, for the purposes of this handbook, estimates with CVs of more than 15 percent are considered cause for caution when interpreting patterns in the data. The choice of a CV level threshold that distinguishes a reliable estimate from an unreliable estimate will vary by application. While CVs for 3- and 5-year estimates would be generally lower than that of the 1-year estimate, one could go with the most recent 1-year estimate if its CV were 15 percent or lower. The illustrations that follow will highlight the computation and use of CVs.

In Figure 1, households receiving public assistance income are shown using 1-, 3-, and 5-year estimates. The period of time represented by each estimate is shown by the dotted lines. While all three estimates share the same final year of data collection, they differ in other important ways. The 1-year (2005) estimate is based on interviews conducted throughout calendar year 2005. The 3-year (2003–2005) estimate is based on the same sample interviews used to produce the three most current 1-year estimates (2003, 2004, and 2005). The same can be said of the 5-year estimates (2001–2005), which include the sample interviews used to produce the five 1-year estimates in 2001, 2002, 2003, 2004, and 2005. This is why the 1-year estimate is preferred when attempting to get a "current" estimate, as it does not draw in data from earlier years. Still, if one looks at the CVs shown at the bottom of Figure 1, it is clear that the 1-year estimate has a higher CV than either the 3- or 5-year estimates. Nevertheless, since the CV for the 1-year estimate from 2005 is under the 15 percent cutoff, we choose to go with that estimate.

What Are Margins of Error and Coefficients of Variation?

Since the estimates of characteristics from the ACS are based on a sample, data are published with margins of error (MOEs) for every estimate. ACS MOEs are based on a 90-percent confidence level. MOEs give users an idea of how reliable, or precise, estimates actually are. For example, there is an ACS estimate of 43,527 kindergarteners in Utah, with an MOE of plus or minus 2,834. The MOE tells us if we had the time and the dollars to create the same estimate of kindergartners several thousand times, from several thousand samples, that 90 percent of the estimates of kindergarteners in Utah would be between 40,693 (43,527–2,834) and 46,361 (43,527+2,834), a fairly precise estimate.

In more technical terms, MOEs provide an idea of how much variability (i.e., sampling error) is associated with the estimate. As an MOE gets larger, relative to the size of an estimate, the estimate becomes less reliable. A measure called the "coefficient of variation" (CV) can also be used to discern the level of reliability of an estimate. This measure is constructed in two steps:

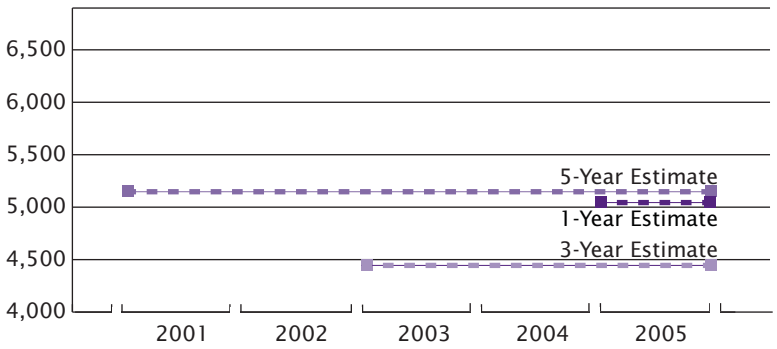
a) Calculate the standard error: SE = MOE/1.645
b) Calculate the Coefficient of Variation: CV = SE/Estimate * 100

With respect to the above example, the standard error associated with the estimate for kindergarteners in Utah is 2,834/1.645 or 1,723. The CV equals 1,723/43,527 * 100, or 4.0 percent. This means that one standard error is 4 percent of the estimate, a fairly low level of variability that indicates that the estimate is reliable. For some applications, CVs that are in excess of 15 percent should be a cause for concern.

With respect to tables provided by the ACS, when the Census Bureau deems that CVs are too high for a tabulation, the entire table may be blanked-out or suppressed. This usually occurs because the data are too sparsely distributed in the cells of a detailed table, which will cause some estimates in a data profile and entire detailed and subject tables to be suppressed. In general, tables based on 5-year estimates are not subject to suppression.

For more information on CVs and MOEs, see Appendix 3.

Figure 1. **Households on Public Assistance, Using 1-Year, 3-Year, and 5-Year ACS Data Anchored on 2005**



Source: U.S. Census Bureau, American Community Survey. Multiyear Estimates Study.

Estimates for 2005 Year of Interest			
	Estimate	MOE	CV
5 Year (2001–2005)	5,091	564	6.7
3 Year (2003–2005)	4,454	760	10.4
1 Year (2005)	5,023	1,174	14.2

While 15 percent is a good cut-off, it is important to consider the application of the data in order to determine whether a specific CV is reasonable. If, for example, a data user is using public assistance data to produce a very general portrait of an area, estimates with CVs beyond 15 percent may be acceptable. If, however, the data are being used to apply for a federal grant, where estimates must be precise, then estimates with lower CVs may be needed. If the data user is working with geographies of under 20,000 people, then there is no recourse but to use the most current 5-year estimate to represent the attributes of an area.

In all cases, it is important for data users who want the most current data to "anchor" estimates on the final time point and then if the option exists, to determine which option best optimizes the trade-off between data that are current and data that are reliable. For those seeking to use the most current estimates, anchoring on the final year makes sense for several reasons. First, the release of estimates is "pinned" to the population threshold of an area at the latest time point, thus providing the most current data. Second, such comparisons are appropriate because when an area changes its geographic boundaries, the geography for the estimate will be current as of the final year of the estimate. Finally, any economic items will be adjusted according to differences with the Consumer Price Index for the last year of the estimate.

Choosing an Estimate From the Recent Past

While data users usually want the most current information available, sometimes data from the more recent past may be required. A new political administration, for example, may want to know the number of public assistance recipients at the start of the prior administration, say in 2002.

Once again, assume that 1-, 3-, and 5-year estimates are available (Figure 2). In this example the 1-year estimate is for 2002, and the appropriate multiyear estimates are now shown as "centered" on 2002, the year of interest. Thus, the 3-year period estimate would be for the years 2001–2003, while the 5-year estimate would be for the years 2000–2004.[4] While the multiyear estimates are not designed to be estimates of the middle year of the series, such comparisons can be useful to see if all available data are telling you the same story. This "centering" on 2002 stands in contrast to "current" estimates in the earlier section that were anchored on the final year.

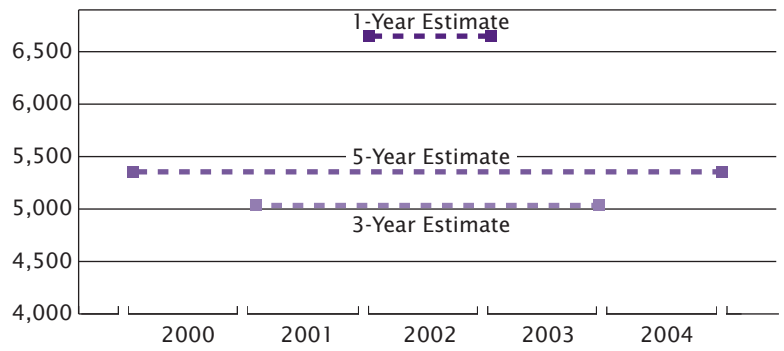
The ideal estimate would be the 1-year estimate, since 2002 is the year of interest. But given its CV of nearly

[4] It is important to note that multiyear estimates were only available for selected test counties prior to 2008. Beginning in December 2008, 3-year estimates will become available for a much broader set of specific areas throughout the United States, followed by 5-year estimates in 2010.

16 percent, one needs to consider the 3- and 5-year estimates. Both the 5-year estimate (5,356) and the 3-year estimate (5,029) shown in Figure 2 have CVs below the 15 percent cutoff. It would be preferable to use the 3-year estimate because it adds only 2 years of additional data to the 1-year estimate. For more information on single- and multiyear estimates and choosing between them, see Appendix 1.

The sections that follow provide three case studies that address a composite of issues faced by state and local governments. The first two case studies use a single set of ACS estimates to characterize an area, while the third case study examines two sets of estimates to gauge change over time. The data analysis involved in each case study offers principles and guideposts for using the ACS. Each case study embodies different aspects of ACS use, with an emphasis on applications that are most pertinent for data users in state and local government. What these case studies show is that effective use of the ACS often requires trade-offs involving the timeliness of the data, the sampling variability, and the geographic fit. The goal was to achieve the best fit between the questions at hand and the available data. Hopefully, an examination of this process will allow data users to develop a sense of what responsible use of the ACS entails.

Figure 2. **Households on Public Assistance, Using 1-Year, 3-Year, and 5-Year ACS Data Centered on 2002**



Source: U.S. Census Bureau, American Community Survey. Multiyear Estimates Study.

Estimates for 2002 Year of Interest			
	Estimate	MOE	CV
5 Year (2000–2004)	5,356	638	7.2
3 Year (2001–2003)	5,029	779	9.4
1 Year (2002)	6,642	1,742	15.9

Case Studies

Distribution of Funds to Senior Centers

Overview

The Department for the Aging (DFTA) in New York County, New York, is working to enhance the quality of life of the elderly in a cost-effective manner. It maintains a network of neighborhood-based senior centers across the county that provide meals and recreational activities, as well as education, health, housing, and social services under the auspices of federal-, state-, and county-sponsored programs. A utilization study recently revealed that use of the centers' resources was unbalanced, with some centers, especially in wealthier areas, experiencing chronic underutilization. A policy decision then followed to tilt the allocation of the annual budget to a few centers located in communities with larger numbers of the elderly who are poor or near poor. This move was seen as a precursor to future allocation efforts, since the local planning department has projected a huge increase in the elderly population over the next 20 years, when cost-effective allocation of limited resources will be even more of a challenge.

DFTA sought to put a program in place that would identify those communities that contained the largest numbers of seniors who are poor or near poor. While the numbers for the entire county were important, this initiative required good data for subcounty areas. The analyst in charge of this initiative was familiar with issues affecting the elderly and had some exposure to census data, mostly via the Census Bureau Web site and through contact with the State Data Center.[5] After talking with the Data Center and searching the Census Bureau's Web site, she concluded that the best source of data for the task at hand was the 2006 ACS. DFTA administrators, however, were a bit reluctant, having heard that the ACS was derived from a small sample, compared with the size of the decennial census sample. While they were eager to get an up-to-date look at a rapidly aging population, they wondered whether they should not just turn to data from the Census 2000 long form. The agency's administrators requested the analyst to present a case for the ACS that conclusively demonstrated how the allocation of resources was best served using these data.

Strategy and Data Sources

The analyst decided that to identify the poor or near poor, the focus should be on poverty estimates, as they are determined by using both income and family/household size. In the American Factfinder (AFF), the analyst located Table C17024 that shows the ratio of income to poverty level for broad age groups.[6] A portion of this table is shown in Figure 3. Refer to the text box for step-by-step instructions of how the analyst found this information.

Step-by-Step Instructions for Case Study 1

Start at <www.census.gov>.

1. Select American FactFinder.
2. Under "Getting Detailed Data," select American Community Survey—get data.
3. ***Choose your data set.*** From the list of data sets, select the 2006 American Community Survey.
4. ***Choose your data product.*** From the list of data products on the right, select detailed tables. The detailed tables contain the greatest variety of information for many topics so it makes sense to start here to search for information on poverty.
5. ***Choose your geography of interest.*** From a series of drop-down lists, select the geographic area type, the state, and then the specific area. For this example select county, New York, and then the specific county within New York—New York County. Click "Add" then "Next."
6. ***Choose your tables.*** To search for information on poverty use the keyword selection method and enter "poverty." Reviewing the list of possible poverty tables, there are two versions of a table on "Age by Ratio of Income to Poverty Level in the Past 12 Months"—B17024 and C17024. Keep in mind that the "B" version has greater detail. It's best to start with the C version—if it has the detail you need, select it. For this example, we are interested in seniors and the age categories in the C table are perfect. Select C17024, click on "Add" then "Show result."
7. ***Consider additional geographies.*** To look at additional subcounty data, select another geographic area type (Public Use Microdata Area), the state (New York), and then select specific PUMAs (PUMAs 03801, 03802, 03803 . . . 03810 are the 10 PUMAs in New York County). Click on "Add" then "Next."

[5] The State Data Center program is a partnership program with the Census Bureau. For more information, see <http://www.census.gov/sdc/www/>.

[6] For more information on the Census Bureau's American FactFinder, go to <http://factfinder.census.gov>.

Figure 3. **Example of Poverty Data for PUMAs in New York County, New York**

U.S. Census Bureau

American FactFinder

Main | Search | Feedback | FAQs | Glossary | Site Map | Help

Detailed Tables

You are here: Main ▸ Data Sets ▸ Data Sets with Detailed Tables ▸ Geography ▸ Tables ▸ **Results**

Use the links above to change your results | Options | Print / Download | Related Items

C17024. AGE BY RATIO OF INCOME TO POVERTY LEVEL IN THE PAST 12 MONTHS - Universe: POPULATION FOR WHOM POVERTY STATUS IS DETERMINED
Data Set: 2006 American Community Survey
Survey: American Community Survey

NOTE. For information on confidentiality protection, sampling error, nonsampling error, and definitions, see Survey Methodology.

View the noncollapsed version of this table.

	PUMA 03801, New York		PUMA 03802, New York		PUMA 03803, New York		PUMA 03804, New York		PUMA 03805, New York	
	Estimate	Margin of Error	Estimate	Margin of Error	Estimate	Margin of Error	Estimate	Margin of Error	Estimate	Margin of Error
Total:	204,457	+/-9,960	111,865	+/-7,939	116,841	+/-7,502	132,102	+/-9,362	211,990	+/-8,918
Under 18 years:	45,605	+/-4,484	20,681	+/-3,960	30,288	+/-3,913	36,382	+/-5,089	33,665	+/-3,894
Under 0.50	7,214	+/-2,795	4,195	+/-2,230	5,475	+/-2,047	5,899	+/-2,203	691	+/-608
0.50 to 0.99	12,522	+/-2,971	3,986	+/-1,631	4,733	+/-1,678	10,981	+/-3,244	166	+/-271
1.00 to 1.24	3,471	+/-1,718	426	+/-332	2,665	+/-1,295	4,633	+/-2,922	258	+/-271
1.25 to 1.99	7,700	+/-2,111	2,789	+/-1,129	5,225	+/-2,008	5,645	+/-1,926	846	+/-754
2.00 to 2.99	9,546	+/-2,327	2,699	+/-1,288	5,000	+/-2,042	3,917	+/-1,666	1,749	+/-1,078
3.00 to 3.99	1,939	+/-868	2,268	+/-1,128	1,926	+/-1,176	2,289	+/-1,196	1,229	+/-535
4.00 to 4.99	1,067	+/-742	525	+/-368	2,472	+/-1,741	1,568	+/-1,037	2,610	+/-1,753
5.00 and over	2,146	+/-752	3,793	+/-1,564	2,792	+/-1,155	1,450	+/-674	26,116	+/-3,439
18 to 64 years	134,029	+/-7,149	75,215	+/-5,370	74,926	+/-5,375	82,723	+/-6,093	149,237	+/-6,793
Under 0.50	10,938	+/-2,240	9,529	+/-2,105	8,901	+/-1,781	12,426	+/-2,074	4,754	+/-1,292
0.50 to 0.99	21,445	+/-3,637	7,687	+/-2,052	10,447	+/-2,263	15,518	+/-2,942	2,623	+/-1,043
1.00 to 1.24	7,370	+/-2,558	3,442	+/-1,325	6,071	+/-2,067	7,530	+/-3,459	2,132	+/-1,128
1.25 to 1.99	21,030	+/-3,628	16,265	+/-3,980	11,202	+/-3,419	9,873	+/-2,228	5,348	+/-1,853
2.00 to 2.99	27,693	+/-4,682	10,771	+/-2,834	10,678	+/-2,319	9,267	+/-2,529	7,780	+/-1,855
3.00 to 3.99	15,703	+/-3,500	5,592	+/-1,737	8,151	+/-2,062	9,703	+/-2,854	10,477	+/-2,213
4.00 to 4.99	10,441	+/-3,218	4,947	+/-1,738	7,073	+/-2,172	8,192	+/-2,380	12,439	+/-2,243
5.00 and over	19,409	+/-3,062	16,982	+/-2,986	12,403	+/-2,307	10,214	+/-2,053	103,684	+/-5,652
65 years and over:	24,823	+/-2,440	15,969	+/-1,899	11,627	+/-1,577	12,997	+/-1,927	29,088	+/-2,704
Under 0.50	510	+/-489	1,448	+/-705	431	+/-298	469	+/-299	804	+/-494
0.50 to 0.99	4,761	+/-1,302	3,711	+/-1,027	3,517	+/-921	3,443	+/-828	1,178	+/-694
1.00 to 1.24	949	+/-378	933	+/-651	1,077	+/-484	1,895	+/-864	1,302	+/-700
1.25 to 1.99	4,431	+/-1,564	3,595	+/-1,128	2,414	+/-703	3,113	+/-1,242	1,742	+/-577
2.00 to 2.99	6,375	+/-1,899	1,494	+/-860	1,594	+/-851	765	+/-350	2,243	+/-786
3.00 to 3.99	3,283	+/-1,402	1,012	+/-706	546	+/-360	863	+/-454	3,048	+/-909
4.00 to 4.99	1,499	+/-802	608	+/-533	372	+/-306	628	+/-515	2,369	+/-843
5.00 and over	3,015	+/-1,047	3,168	+/-1,233	1,676	+/-629	1,821	+/-1,198	16,402	+/-2,223

Source: U.S. Census Bureau, 2006 American Community Survey

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

Using Table C17024 with its age groupings, poverty ratio groupings, estimates, and MOEs for the county and its PUMAs, it was possible to carry out a detailed analysis of the distribution of economically distressed elderly people. This table has major advantages. First, the data are available for different age groups, which make it possible to separate out the elderly population based on the agency's definition of the elderly as people 65 years and over. Second, the various ratios provide a great deal of flexibility in identifying the poor and near poor. Third, since this is a large county, 1-year data are available for the county overall and for its 10 PUMAs, making these data well suited for the application. As mentioned earlier, PUMAs are areas with at least 100,000 residents in contiguous census tracts. The tables also include MOEs for each estimate, which permit the analyst to calculate coefficients of variation and provide a detailed assessment of data reliability. (See the text box titled "What Are Margins of Error and Coefficients of Variation?")

The analyst used options available in the AFF to download Table C17024 into a spreadsheet and was able to isolate the specific data of interest—the data for people aged 65 and over. Examining the poverty categories available in Table C17024, the analyst determined that a ratio of income to poverty level of less than 1.0 identifies the population below the poverty level. Persons

Table 2. **Persons 65 and Over Who Are Below Twice the Poverty Level in a Selected County and Its Component PUMAs, 2006**

	Individual PUMAs				Individual and Merged PUMAs				
	Estimate	MOE	CV		Estimate	MOE	CV	Persons per Acre
Selected County	81,877	5,061	3.8	Selected County	81,877	5,061	3.8	5.6
PUMA 3801	10,651	2,127	12.1	PUMA 3801	10,651	2,127	12.1	5.4
PUMA 3802	9,687	1,802	11.3	PUMA 3802	9,687	1,802	11.3	11.3
PUMA 3803	7,439	1,291	10.5	PUMA 3803	7,439	1,291	10.5	8.0
PUMA 3804	8.920	1,750	11.9	PUMA 3804	8,920	1,750	11.9	6.0
PUMA 3805	5,026	1,244	15.1	PUMA 3805 & 3806	14,185	2,248	9.6	4.1
PUMA 3806	9,159	1,873	12.4					
PUMA 3807	5,819	1,334	13.9					
PUMA 3808	4,761	1,068	13.6	PUMA 3808	4,761	1,068	13.6	4.5
PUMA 3809	15,579	2,326	9.1	PUMA 3809	15,579	2,326	9.1	14.0
PUMA 3810	4,836	1,551	19.5	PUMA 3807 & 3810	10,655	2,046	11.7	2.8

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

with a ratio of less than 2.0 are persons living below twice the poverty level. The analyst decided that for her analysis elderly people with incomes below twice the poverty level would be considered poor or near poor. To obtain data for those below twice the poverty level, the estimates of people in four income-to-poverty ratios had to be summed: under 0.50, 0.50 to 0.99, 1.00 to 1.24, and 1.25 to 1.99. These combined estimates are shown in Table 2.

Analysis and Findings

While summing estimates of poverty across four income-to-poverty ratios gives one the estimate for the combined category, summing MOEs does not produce the correct MOE. The analyst found the formula that she needed to produce the right MOE for a derived estimate in Appendix 3.

Using this formula for each PUMA, the analyst calculated the MOE for the combined category of people age 65 and over with an income under twice the poverty level. Using the data in this table, she had to determine whether the estimates were reliable enough to satisfy DFTA administrators. The MOEs, however, seemed to vary with the estimates, with large estimates producing large MOEs. With further sleuthing, the analyst decided to calculate CVs, which are standardized indicators of reliability, thus allowing comparisons of reliability across estimates. (See the text box titled "What Are Margins of Error and Coefficients of Variation?") While thresholds for acceptable CVs have always been a subject of debate, she decided to use a threshold of 15 percent; any estimate with a CV over 15 percent would be considered unreliable.

Two out of the 10 PUMAs—PUMA 3805 and PUMA 3810—had CVs that were over the 15 percent threshold. But with so many reliable estimates, the analyst did not want to abandon the current ACS data for outdated decennial data. So, the analyst decided to merge each of the PUMAs with a CV over the threshold with an adjacent PUMA. PUMA 3805 was combined with the adjacent PUMA 3806, while PUMA 3810 was combined with neighboring PUMA 3807. This geographic aggregation increased the sample size in the merged PUMAs, resulting in a decline in their CVs to under the 15 percent threshold (right half of Table 2). This assured DFTA administrators that the subcounty estimates of the elderly population at or near poverty were reliable. But the overall number of geographic units available for analysis declined from the original 10 units to 8 units (from 10 PUMAs to 6 PUMAs plus 2 areas, each with 2 merged PUMAs).

It became clear that the county was far from uniform in its distribution of the elderly poor and near poor. One PUMA (3809) had more than three times the number of another PUMA (3808). The administrators wanted to know whether these estimates could be rank ordered by the size of the target population. The analyst warned the administrators not to focus on strict rank-

ings or small differences between the PUMA estimates since the sample was not capable of making such fine distinctions. To help avoid "over-analysis" of fine differences between two estimates, the analyst suggested a rule of thumb: If the first estimate +/− its MOE overlapped the second estimate +/− its MOE, the difference between the two estimates would not be considered to be statistically significant. Using this rule, the size of the elderly poor and near poor in PUMA 3809 (15,579 +/− 2,326 = 13,253 to 17,905) was not significantly different from the merged PUMAs 3805/3806 (14,185 +/− 2,248 = 11,937 to 16,433). Similarly, the estimate of the elderly poor for PUMA 3801 was not statistically different from PUMAs 3807/3810 or PUMA 3802. Thus, there were many PUMAs that had large target populations that were similar in size. While this rule of thumb is useful, it is not a definitive test of statistical significance.[7]

The analyst was also aware that the merged PUMAs resulted in a larger target population. So for each PUMA, as well as for the merged PUMAs, she calculated the population density of the targeted group, defined as the population per acre of the poor and near poor. This helped her gauge the numerical strength of the elderly poor and near poor in the context of the size of the PUMAs, especially the merged PUMAs.

Taking into account current utilization of senior centers in PUMAs and the size and density of the elderly poor and near poor population, the analyst decided to recommend focused funding for senior centers in two PUMAs. The first PUMA recommended was PUMA 3809, since it had both the largest target population and highest population density. She decided against recommending the merged PUMAs 3805/3806 and 3807/3810 though each had a large population of elderly poor and near poor residents, since the density of this population was among the lowest in the county and utilization of senior centers was below average. Instead, the second PUMA recommended was PUMA 3802, with nearly 9,700 elderly poor and near poor residents and the second highest population density in the county.

One other option the analyst could have considered to make estimates more reliable was to employ estimates using data at the PUMA level for 3-year and/or 5-year periods. The analyst was able to use 1-year estimates because the county of interest was very large. In fact, for most small governments (whose areas have fewer than 65,000 people), only estimates for 3 and/or 5 years will be available. In general, considerable increases in reliability over 1-year estimates can be achieved by using the 3- and 5-year period estimates, but this comes at a cost. As each additional year is added to the estimate, it becomes less current, covering a larger period of time. Along with the use of 5-year estimates at the PUMA level, the analyst could also have looked at 5-year census tract data to get a better sense of the population density of the elderly poor and near poor in the "catchment" area of each senior center.

Summary: What Have We Learned?

Effective use of the ACS can be said to represent a "healthy tension" between the data required to answer a question and the statistical limits of those data. **There is always more to learn through the use of more detailed table categories, but this comes with a compromise in the form of lower levels of reliability**. All ACS applications require a balance between content detail and reliability. The use of standard errors, as seen in the MOEs, is essential information that enables a data user to optimize this relationship. Using the decision-making guidelines presented in the previous application, it is possible for data users to make the most of the data, while still maintaining a sense that the numbers are within an acceptable level of reliability. Decisions become difficult when they require a loss of content, as with the decision to combine PUMAs, whose level of reliability was questionable.

Collapsing of geographic areas is one of several ways of enhancing reliability; collapsing data categories is another. Both techniques were adopted here, though the collapsing of the income-to-poverty ratios was required conceptually, rather than being employed as a means to improve reliability. Once the 3- and 5-year averages become available, it will be possible to obtain more reliable data for individual PUMAs.

Finally, the most important point to understand is that the path chosen is not solely a statistical decision but one also based on the question at hand and the purpose of the application. A firm idea of what one is trying to accomplish with the data is the critical prerequisite for using the ACS.

Fire Prevention Education for Those With Limited English Proficiency

Overview

In March of 2007, a nighttime fire swept through a small apartment building in a New York City neighborhood, killing one adult and nine children from the African country of Mali. Of the factors that contributed to the lethal nature of the fire, the absence of batteries in the building's two smoke alarms, was the most troubling for local officials. This tragedy highlighted the fact that more needed to be done to promote fire safety awareness among recent immigrants, especially those who spoke little or no English.

[7] For exact methods of significance testing see Appendix 4, "Making Comparisons."

Subsequent to this tragedy, a task force was setup in Queens County to initiate a fire safety campaign to raise public awareness about the importance of maintaining functioning smoke alarms. This campaign was to be targeted to residents who were not fluent in English. The campaign would consist of a series of television, radio, print, and Internet ads in the major languages spoken by those not proficient in English.

Strategy and Data Sources

The first order of business was to find out the number of people who had limited English-speaking skills and the languages they spoke. These data would be needed for the overall county and, for the purposes of grassroots communication efforts, for subcounty areas.

Using the Census Bureau's AFF, 2006 ACS data can be accessed to find the latest information on languages spoken and English-speaking ability. Starting with the data profiles for Queens County, New York, the 2006 ACS turned up a generalized list of languages and English language ability for the population 5 years and over, but the only specific language not subsumed into a larger linguistic grouping was Spanish. A portion of this data profile is shown in Figure 4. Further research using detailed tables for the same geographic area revealed Table B16001, which appeared to be a perfect fit. Part of this table is shown in Figure 5. It shows language spoken at home by ability to speak English for the population 5 years and over. Also, the table is very detailed, with 38 languages or language groupings.

Figure 6 provides a general summary of these data. It shows that the estimate of the population in Queens County reporting that they speak English less than "very well" numbered over 620,000, and Spanish speakers were the number one group, accounting for over 43 percent of the total. Chinese (17 percent) and Korean (7 percent) rounded out the top three languages, which together accounted for about two-thirds of the population in the county with limited English-speaking abilities.

As these data compilation efforts were underway, funding for the task force's outreach efforts was cut and the task force decided to limit its outreach to the most hard-to-reach segment of this population—residents who had both limited English-speaking skills and less than a high school education. Since a table that crossed language abilities by educational attainment was not available in the AFF, the task force turned to the 2006 ACS PUMS file, which allows a user to create detailed tabulations down to the PUMA level. (See the text box titled "What Is the Public Use Microdata Sample?") The tabulations one can derive from the PUMS are limited only by the number of records in the file that meet the criteria desired. For details on how to obtain PUMS records, refer to the handbook in this series for PUMS data users.

PUMS records for the county were first identified, followed by the records for people not proficient in English. These were defined as those who spoke English less than "very well."[8] The data were further

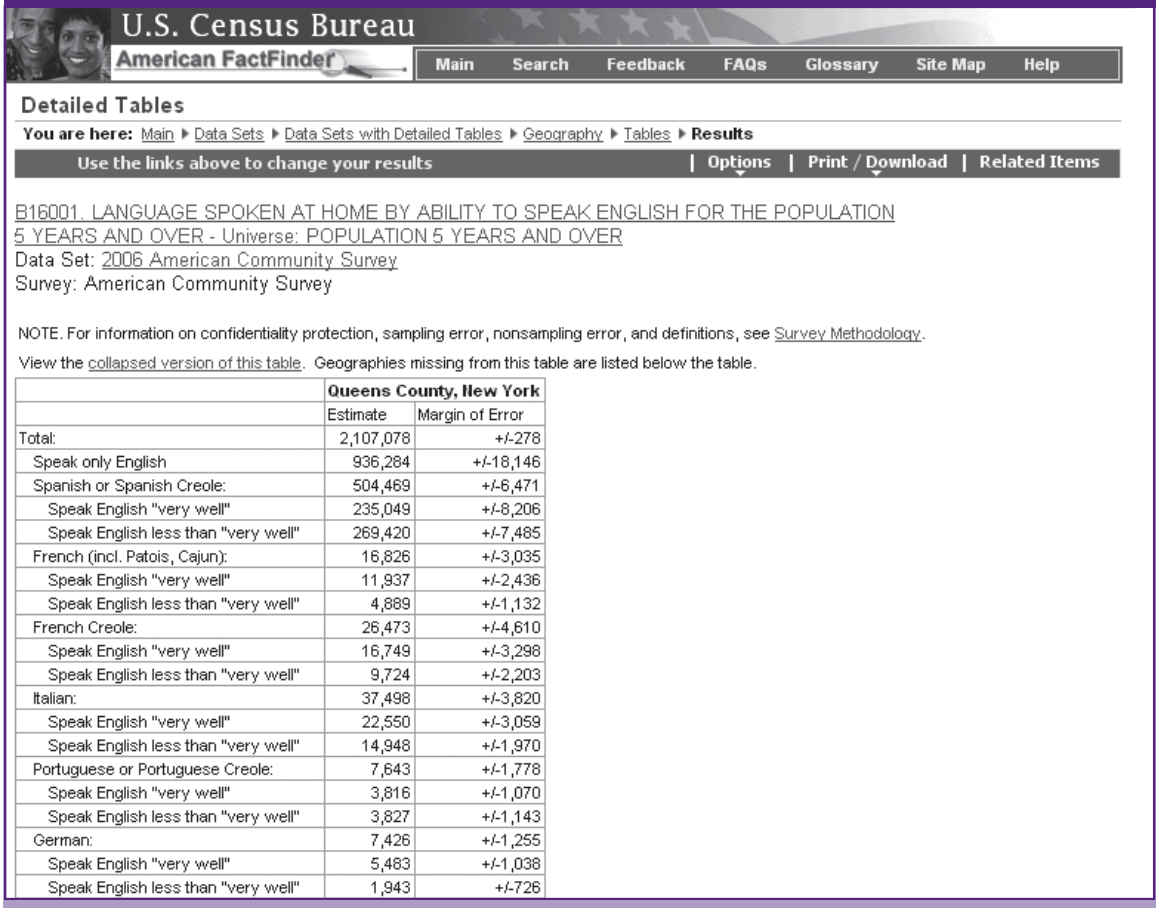
Figure 4. **Information Available on Language Spoken at Home From ACS Data Profile**

Selected Social Characteristics in the United States: 2006	Estimate	Margin of Error
Northern America	4,022	+/-1,120
LANGUAGE SPOKEN AT HOME		
Population 5 years and over	**2,107,078**	**+/-278**
English only	936,284	+/-18,146
Language other than English	1,170,794	+/-18,190
Speak English less than "very well"	620,238	+/-13,075
Spanish	504,469	+/-6,471
Speak English less than "very well"	269,420	+/-7,485
Other Indo-European languages	349,138	+/-13,218
Speak English less than "very well"	160,015	+/-9,086
Asian and Pacific Islander languages	280,467	+/-9,838
Speak English less than "very well"	176,248	+/-8,092
Other languages	36,720	+/-7,092
Speak English less than "very well"	14,555	+/-3,726

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

[8] The Census Bureau uses four categories of English language ability for households that report speaking a language other than English at home: those that speak English "very well," "well," "not well," and "not at all." Those who speak English well, not well, or not at all are considered to have limited English proficiency.

Figure 5. **Information Available on Language Spoken at Home From ACS Detailed Table, Queens County, New York, 2006 American Community Survey**



U.S. Census Bureau
American FactFinder
Main | Search | Feedback | FAQs | Glossary | Site Map | Help

Detailed Tables

You are here: Main ▸ Data Sets ▸ Data Sets with Detailed Tables ▸ Geography ▸ Tables ▸ Results

Use the links above to change your results | Options | Print / Download | Related Items

B16001. LANGUAGE SPOKEN AT HOME BY ABILITY TO SPEAK ENGLISH FOR THE POPULATION 5 YEARS AND OVER - Universe: POPULATION 5 YEARS AND OVER
Data Set: 2006 American Community Survey
Survey: American Community Survey

NOTE. For information on confidentiality protection, sampling error, nonsampling error, and definitions, see Survey Methodology.
View the collapsed version of this table. Geographies missing from this table are listed below the table.

	Queens County, New York	
	Estimate	Margin of Error
Total:	2,107,078	+/-278
Speak only English	936,284	+/-18,146
Spanish or Spanish Creole:	504,469	+/-6,471
Speak English "very well"	235,049	+/-8,206
Speak English less than "very well"	269,420	+/-7,485
French (incl. Patois, Cajun):	16,826	+/-3,035
Speak English "very well"	11,937	+/-2,436
Speak English less than "very well"	4,889	+/-1,132
French Creole:	26,473	+/-4,610
Speak English "very well"	16,749	+/-3,298
Speak English less than "very well"	9,724	+/-2,203
Italian:	37,498	+/-3,820
Speak English "very well"	22,550	+/-3,059
Speak English less than "very well"	14,948	+/-1,970
Portuguese or Portuguese Creole:	7,643	+/-1,778
Speak English "very well"	3,816	+/-1,070
Speak English less than "very well"	3,827	+/-1,143
German:	7,426	+/-1,255
Speak English "very well"	5,483	+/-1,038
Speak English less than "very well"	1,943	+/-726

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

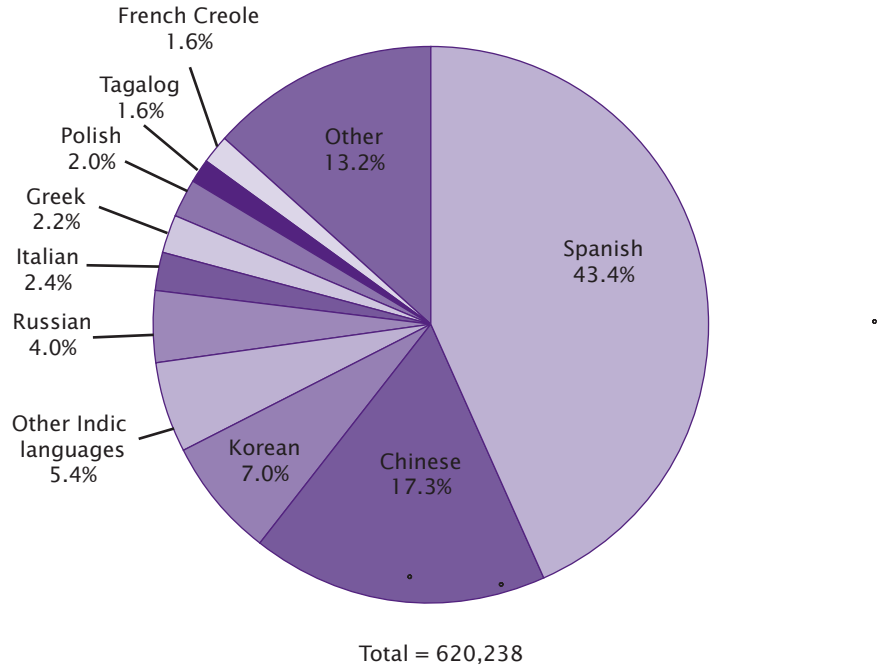
limited to people 25 years and over who had less than a high school education. The age cutoff was important since most people have completed their high school education by age 25. Once the population was limited to those 25 years and over without a high school diploma and not proficient in English, it was possible to ascertain the languages they spoke at home.

Analysis and Findings

While the estimate of the total population in Queens County with limited English-speaking abilities was about 620,000, Figure 7 shows that the estimate of the population 25 years and over without a high school diploma with limited English-speaking abilities was only about 185,000. Adding educational attainment successfully limited the task force target population. The subset that spoke Spanish numbered over 98,000, accounting for about 53 percent. (It's worth noting that Spanish speakers are overrepresented in this subgroup, as they accounted for 43 percent of the total population with limited English-speaking abilities.) Chinese speakers accounted for about 16 percent, while Italian speakers accounted for about 5 percent. Thus, the top 3 languages for the population without a high school diploma were slightly different than for the total population. These three groups accounted for nearly three-quarters of the population 25 years and over without a high school diploma and not proficient in English. If it were important to be sure which language groups were most prevalent, it would have been best to include the margins of error or to conduct statistical testing. For this initial assessment, the task force was only interested in a broad picture of the language needs.

Next, similar estimates were produced for each PUMA in Queens County. Table 3 displays these estimates along with the associated MOEs and CVs. As described in the first case study, the MOEs are provided with the tables but the CVs must be calculated using the simple formula provided in Appendix 3. The task force chose to apply a 15 percent CV threshold to identify reliable estimates for targeting programs. Using this definition, 6 of the 14 PUMAs had reliable estimates of the population age 25 and over with limited English-speaking abilities and less than a high school

Figure 6. **Top Languages Spoken at Home in a Selected County for the Population 5 Years and Over Who Are Limited English Proficient, 2006**



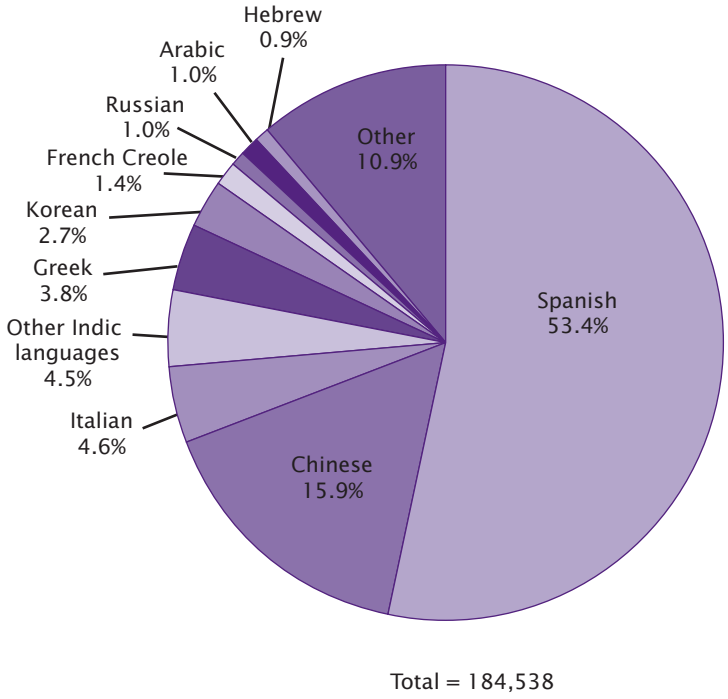
Source: U.S. Census Bureau, American FactFinder, accessed at <http://factfinder.census.gov>.

education. For the proportion that were Spanish speaking, 6 of the 14 PUMAs had reliable estimates that could be used for targeting purposes, while just 1 PUMA for Chinese speakers had a reliable estimate. The countywide figures for the population that spoke Italian, Greek, and Korean were not reliable and neither were the PUMA estimates. Nevertheless, there was enough information by county and PUMA for the total limited English-speaking, less than high school diploma population, as well as for the subset of Spanish speakers, to guide the taskforce in its targeting efforts.

Summary: What Have We Learned?

This example demonstrates how to use three different data products to gather information about language spoken at home and English-speaking ability. The data profiles provided the least detailed data, which can be sufficient for some needs but were not useful for this application. The detailed tables provided greater detail, but not the optimal information. The PUMS file offered a useful alternative for local government data users when the detail needed to successfully implement a program is beyond what is available in the AFF. This flexibility comes at a cost in the form of smaller samples from which to derive estimates. For large counties, PUMS data provide rich opportunities to explore myriad tabulations that may be extremely useful in setting up and carrying out programs. The desire for additional detail in the tabulations needs to be balanced against the decrease in reliability, especially at the PUMA level. The reliability of the estimates shown in Table 3 are reduced due to the partitioning of the population into increasingly smaller subgroups—the population age 25 and over that reported speaking a specific language at home, reported speaking English less than "very well," and reported not having a high school diploma. Reliability, however, can be enhanced by using 3- and 5-year PUMS files, when they eventually become available.

Figure 7. **Top Languages Spoken at Home in a Selected County for the Population 25 Years and Over Who Are Limited English Proficient and Have Less Than a High School Education, 2006**



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

Table 3. **Top Languages Spoken at Home in a Selected County and Its PUMAs for the Population 25 Years and Over Who Are Limited English Proficient and Have Less Than a High School Education, 2006**

	Total			Spanish			Chinese*			Italian			Greek			Korean		
	Estimate	MOE	CV	Estimate	MOE	CV	Estimate	MOE	CV	Estimate	MOE	CV	Estimate	MOE	CV	Estimate	MOE	CV
Selected County	184,538	12,127	4.0	98,487	6,314	3.9	29,317	4,626	9.6	8,467	2,648	19.0	6,937	2,407	21.1	4,942	2,043	25.1
PUMA 4101	21,779	4,083	11.4	12,348	2,154	10.6	400	584	88.8	1,665	1,155	42.2	2,847	1,466	31.3	-	-	-
PUMA 4102	28,272	4,559	9.8	24,543	1,676	4.2	2,654	1,445	33.1	108	306	172.2	193	408	128.5	-	-	-
PUMA 4103	22,632	4,152	11.2	3,837	1,663	26.3	11,406	2,216	11.8	1,248	1,012	49.3	1,566	1,125	43.7	2,047	1,271	37.7
PUMA 4104	7,629	2,520	20.1	2,036	1,138	34.0	2,157	1,159	32.7	605	695	69.8	856	812	57.7	932	843	55.0
PUMA 4105	5,789	2,206	23.2	2,237	1,091	29.6	-	-	-	144	349	147.3	-	-	-	-	-	-
PUMA 4106	7,904	2,563	19.7	2,699	1,242	28.0	2,404	1,205	30.5	175	385	133.7	418	586	85.2	269	475	107.3
PUMA 4107	25,422	4,362	10.4	16,949	2,214	7.9	4,295	1,760	24.9	612	720	71.5	154	364	143.7	1,027	925	54.8
PUMA 4108	6,351	2,307	22.1	1,092	886	49.3	1,879	1,072	34.7	-	-	-	-	-	-	-	-	-
PUMA 4109	15,283	3,488	13.9	8,219	1,816	13.4	2,150	1,266	35.8	125	328	159.5	644	732	69.1	464	625	81.9
PUMA 4110	13,300	3,273	15.0	6,596	1,699	15.7	751	784	63.5	2,851	1,394	29.7	-	-	-	-	-	-
PUMA 4111	9,293	2,768	18.1	5,856	1,371	14.2	607	702	70.3	72	249	210.2	-	-	-	-	-	-
PUMA 4112	8,128	2,597	19.4	5,314	1,264	14.5	336	529	95.7	88	275	190.0	-	-	-	72	249	210.2
PUMA 4113	5,644	2,179	23.5	3,168	1,098	21.1	141	345	148.7	774	761	59.8	160	367	139.4	60	227	230.0
PUMA 4114	7,112	2,436	20.8	3,593	1,242	21.0	137	342	151.8	-	-	-	99	291	178.7	71	247	211.5

* Chinese combines the PUMS languages "Chinese," "Mandarin," and "Cantonese."
Source: U.S. Census Bureau, American FactFinder, accessed at <http://factfinder.census.gov>.

Examing Growth in the Foreign-Born Population

Overview

Neighborhood advocacy groups in a city of about 50,000 have been vocal about the effects of a growing foreign-born population on their community. They have deluged the city government with requests for programs to help recent immigrants acclimate to their new environment. With the budget already strained and competing requests from other groups for city services, government leaders need to make tough decisions. In an effort to get a handle on the situation, the city manager has asked a data analyst to establish whether there has indeed been an increase in foreign-born residents in this community.

Strategy and Data Sources

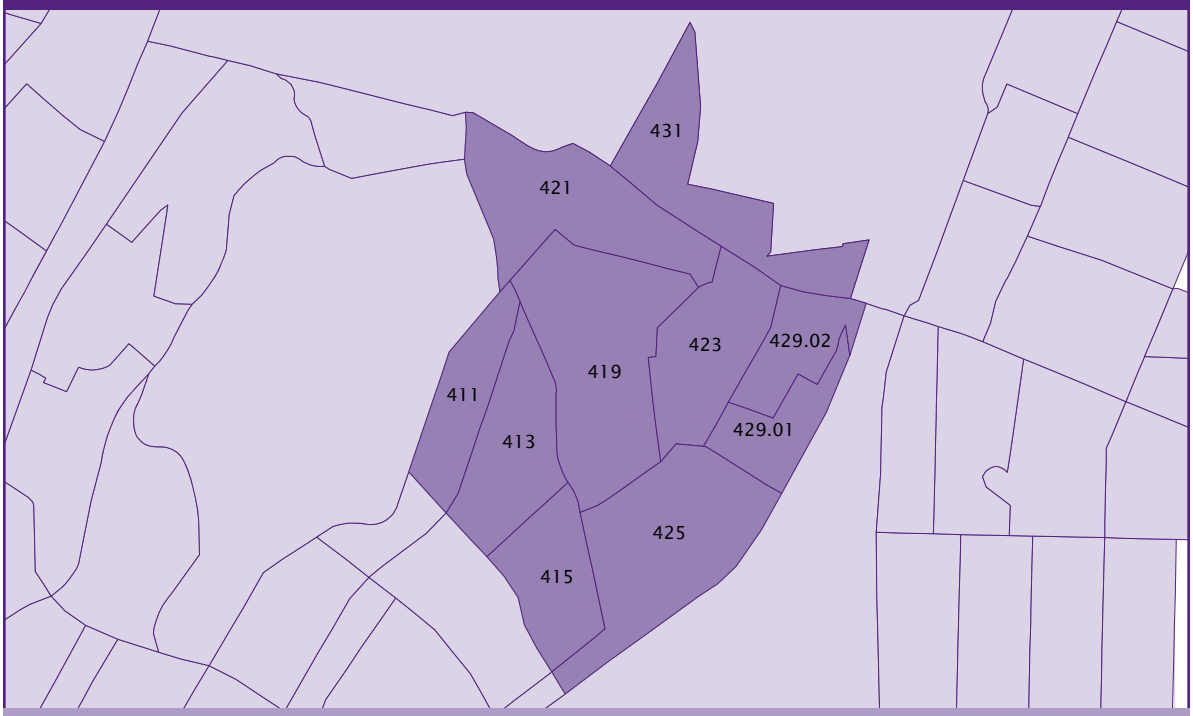
This fictitious community that we will call Bedford, is not a political/administrative entity that would be separately recognized by the Census Bureau in its tabulations, but is an established neighborhood that is recognized by the city. Since the data must be "current," decennial census long-form data for 1990 and 2000 are not useful for this application. The data analyst turns to the ACS for a look at data in the post-2000 period.

Since Bedford is much smaller than the city in which it is contained, city-level data cannot be used as a basis for analysis. The city manager has identified "geographic fit" as the most important dimension of his request, so the application must provide data for a geographic area that would appropriately encompass Bedford. The only relevant data available are 5-year period estimates from the AFF for each of the 10 census tracts that roughly correspond to Bedford (Figure 8). Since the city is located in an ACS test county, three 5-year census tract estimates are available: 1999–2003, 2000–2004, and 2001–2005. Figure 9 displays the time periods covered by each of these three period estimates.

These three 5-year estimates allow three possible comparisons: 1999–2003 vs. 2000–2004, 2000–2004 vs. 2001–2005, and 1999–2003 vs. 2001–2005. The first thing the data analyst has to decide is which comparison best establishes whether there has been an increase in immigrants in Bedford.

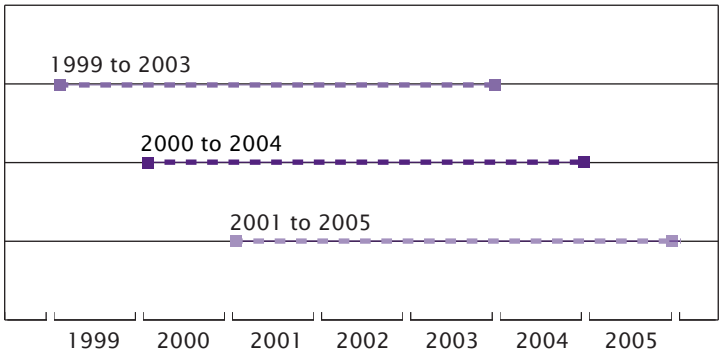
The first two comparisons cover a 6-year period with 4 years of overlap (Table 4). For example, when comparing 1999–2003 with 2000–2004, the four overlapping years are 2000, 2001, 2002, and 2003, and change is solely a function of the difference between year 1

Figure 8. **Census Tracts Comprising Bedford**



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

Figure 9. **5-Year Tract Estimates Available for 1999–2005**



Source: U.S. Census Bureau, American Community Survey.

(1999) and year 6 (2004). When comparing 2000–2004 with 2001–2005, the 4 overlapping years are 2001, 2002, 2003, and 2004, and change is solely a function of the difference between 2000 and 2005. The 4 years of overlap result in a situation where change in the two 5-year period estimates is likely to be statistically significant only if the magnitude of change between year 1 (1999) and year 6 (2004), in the case of the 1999–2003 and 2000–2004 comparison, is relatively large.

The third comparison, 1999–2003 versus 2001–2005, covers 7 years with three overlapping years: 2001, 2002, and 2003. This means that the difference between the 1999–2003 and 2001–2005 estimates is a function of the change between the two nonoverlapping time periods: 1999–2000 and 2004–2005. The data analyst realizes this comparison is likely to prove more useful than the others due to the greater length of the entire comparison period (7 years) and the smaller degree of overlap (3 years). Eventually, as more of these 5-year period estimates become available, it will be possible to compare data for periods with less overlap and, ultimately, two 5-year estimates with no overlap.

Analysis and Findings

Table 5 shows the city, neighborhood, and tract-level estimates for the foreign-born population for the periods 1999–2003 and 2001–2005 and displays the estimates of change between these two time periods. To increase reliability, the data analyst aggregated the 10 census tracts and created an estimate for the entire neighborhood of Bedford. Appendix 3 explains how to calculate margins of error for a derived estimate such as this neighborhood aggregation.

Table 4. **Comparisons of Overlapping 5-Year Estimates, 1999–2005**

	ACS Years						
Estimate Period	1999	2000	2001	2002	2003	2004	2005
4 years of overlap							
1999–2003	(X)	X	X	X	X		
2000–2004		X	X	X	X	(X)	
2000–2004		(X)	X	X	X	X	
2001–2005			X	X	X	X	(X)
3 years of overlap							
1999–2003	(X)	(X)	X	X	X		
2001–2005			X	X	X	(X)	(X)

X – Overlapping Year (X) – Unique Year (Nonoverlapping Year)

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

This handbook has focused on the margin of error (MOE) and coefficient of variation (CV) as tools to evaluate the reliability of ACS estimates. The use of MOEs for assessing the reliability of change over time is more complicated, especially when change is being examined 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 local governments whose only source of data will be 5-year estimates will not want to wait until 2015 to evaluate change (i.e., compare estimates for 2005–2009 with those for 2010–2014). Also, there may be instances where 5-year nonoverlapping estimates are available but a small local government may not want to go back 10 years to examine change, so it chooses a later starting point and some level of overlap (e.g., 2007–2011 and 2010–2014).

Change between these two 5-year estimates involves an evaluation of change in the nonoverlapping years (1999 and 2004, 2000 and 2005). Statistical testing was conducted and results of significant differences are flagged in Table 5. Calculating the difference between these two period estimates is simple, but calculating the MOEs associated with this difference requires the recognition that the estimates include common sample interviews. Refer to Appendix 1 and Appendix 4 for more guidance on making such comparisons, interpreting such comparisons, and calculating MOEs for differences derived from such comparisons.

Walking through the statistical testing for the neighborhood of Bedford, the standard errors (SEs) can be derived directly from the published MOEs. For the 1999–2003 estimate the SE is 967, for the 2001–2005 the SE is 1,083. The value of C is determined as the fraction of overlapping years, here 3 out of 5 or 0.6. Appendix 4 instructs us to approximate the SE of this difference as:

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

$$SE(21{,}824 - 19{,}161) \cong \sqrt{(1-0.6)}\sqrt{1{,}083^2 + 967^2}$$

Thus the SE of the difference value of 2,663 is 918. Appendix 4 explains that the ratio of the estimate of the difference to the SE of that difference gives you a test value that, when compared with a critical value (1.645 for example for a confidence level of 90 percent) allows you to determine if that difference is statistically significant. This means that the user can be certain to a specified degree that the observed difference in the two estimates was not due to chance. In this example the ratio is 2.9. Since it is greater than 1.645 we can be confident (at 90 percent) that the difference of 2,663 is statistically significant.

A look at change over time reveals what looks like substantial increases in several census tracts, but only one tract increase is statistically significant. Change at the neighborhood level shows that the number of foreign-born people has indeed increased by a substantial margin, from 19,161 for the 1999–2003 period to 21,824 for 2001–2005, a statistically significant increase of nearly 2,700 people, or about 14 percent. This compares with the citywide increase of about 8 percent.

Table 5. **Change in the Foreign-Born Population of a Selected City, and the Neighborhood of Bedford and Its Component Census Tracts, 1999–2003 and 2001–2005**

	1999–2003	2001–2005	Change, 1999–2003 to 2001–2005	
	Estimate	Estimate	Number	Percent
Selected City	48,600	52,454	3,854*	7.9*
Bedford	19,161	21,824	2,663*	13.9*
Census Tract 041100	916	1,034	118	12.9
Census Tract 041300	2,778	3,553	775	27.9
Census Tract 041500	2,211	3,104	893*	40.4*
Census Tract 041900	1,718	2,094	376	21.9
Census Tract 042100	2,213	2,335	122	5.5
Census Tract 042300	1,161	1,400	239	20.6
Census Tract 042500	2,331	2,470	139	6.0
Census Tract 042901	902	859	–43	–4.8
Census Tract 042902	1,492	1,771	279	18.7
Census Tract 043100	3,439	3,204	–235	–6.8

*Significant at the 0.10 level.

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

The data analyst concludes that Bedford has indeed experienced a substantial increase in its foreign-born population over this period.

Summary: What Have We Learned?

While medium-sized cities, like the one in this case study, could have 1-, 3-, and 5-year estimates, these data may not reflect what is going on in various neighborhoods. Small areas, such as neighborhoods, that are not political/administrative units will have to rely on aggregating census tract data, which are only available as 5-year estimates. When working with these estimates, it is best to minimize the overlap between periods being compared. In general, estimates with 4 years of overlap should be avoided altogether, in that differences are entirely a function of change between just two single nonoverlapping years. A comparison of estimates with 3 years of overlap is an improvement, but ultimately, as the ACS program continues to move forward, 5-year estimates with little or no overlap should be employed.

Combining census tracts is a good way to produce reliable estimates, as was evidenced by the combination of the set of census tracts that comprised Bedford. However, the data analyst could have elected to combine just a few census tracts within Bedford to see how growth in the immigrant population was distributed within the neighborhood. Indeed, combining census tracts 041300, 041500, and 041900 accounted for about three-quarters of the foreign-born increase in Bedford, which is statistically significant. Such an observation can help local officials target services provided by local groups to specific areas experiencing the largest increases in immigrants.

This case study demonstrates the power of using 5-year estimates as building blocks to create neighborhood statistics, but similar approaches could be used to combine 3-year estimates for areas with populations over 20,000 into larger geographic areas of particular interest. Such aggregations improve the reliability of the estimates.

Conclusion

The American Community Survey (ACS) represents new terrain for data users. Its greatest strength is that it will offer data on an annual basis, but this also results in an array of options that will affect how state and local governments use the data. This handbook illustrates some of the ways that the ACS can be used to achieve its full potential to address issues that are pertinent to state and local governments.

Understanding multiyear period estimates is critical to the proper use of these new data. The concept of a multiyear estimate is, in itself, a big leap for most data users who are used to the April 1 reference point for decennial census data. However, multiyear estimates will be the primary data source for most governments. Of the approximately 53,000 states, counties, cities, urban areas, towns, townships, villages, other minor civil divisions, and census designated places, well over 90 percent will rely on multiyear estimates, with most using 5-year period estimates exclusively. This is because most local governments are small, covering geographic areas with fewer than 20,000 people. Table 1 provides information on the availability of estimates by type of geographic area.

Still, for some geographic areas, such as counties, there will be considerable flexibility in the choice of estimates. One-quarter of all counties—those with at least 65,000 people—will have the full choice of 1-, 3-, and 5-year estimates. These counties account for more than 80 percent of the nation's population. For these areas, and other governmental jurisdictions in the same size class, their choices involve a more complex series of decisions regarding different sets of estimates. For example, the presence of multiple multiyear estimates increases the opportunities for measuring change but adds the burden of deciding which set of estimates makes the most sense to use. In the end, what makes the most sense is a matter of judgment regarding the balance between the period of time covered by an estimate and its level of reliability. The key is to strive to use only reliable estimates, where the time period covered best suits the question at hand.

Good judgment also entails increased attention to a number of data features, some of which have always been with us but have heretofore been largely confined to footnotes and appendixes. The most salient is the heightened importance of sampling variability on whether data are useful "straight out of the box," or whether some of the strategies described herein (e.g., combining geographic areas or data categories) are required to make the data reliable.

Finally, there is the issue of how to use multiyear characterizations of an area to measure change over time. As the ACS moves forward, a series of multiyear estimates for various time intervals will become available. Once successive sets of 3-year and 5-year averages become available with little or no overlap, the value of ACS data will become even more evident. Moreover, as a "feedback loop" develops between the Census Bureau and the data user community, new perspectives and solutions regarding the use of multiyear averages are likely to emerge.

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 MOE to the estimate (for the upper bound) and subtracting the MOE from the estimate (for the lower bound). All published ACS MOE are based on a 90-percent confidence level.

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

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

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

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

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

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

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

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

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

Quality measures. Statistics that provide information about the quality of the ACS data. The ACS releases four different quality measures with the annual data release: 1) initial sample size and final interviews; 2) coverage rates; 3) response rates, and; 4) item allocation rates for all collected variables. The ACS Quality Measures Web site provides these statistics each year. In addition, the coverage rates are also available for males and females separately.

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

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

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

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

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

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

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

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

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

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

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

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

Appendix 1.

Understanding and Using ACS Single-Year and Multiyear Estimates

What Are Single-Year and Multiyear Estimates?

Understanding Period Estimates

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

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

Because ACS estimates provide information about the characteristics of the population and housing for areas over an entire time frame, ACS single-year and multiyear estimates contrast with "point-in-time" estimates, such as those from the decennial census long-form samples or monthly employment estimates from the Current Population Survey (CPS), which are designed to measure characteristics as of a certain date or narrow time period. For example, Census 2000 was designed to measure the characteristics of the population and housing in the United States based upon data collected around April 1, 2000, and thus its data reflect a narrower time frame than ACS data. The monthly CPS collects data for an even narrower time frame, the week containing the 12th of each month.

Implications of Period Estimates

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

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

Table 1. **Percent in Labor Force—Winter Village**

Month											
Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.
20	20	40	60	60	60	60	60	60	50	30	20

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

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

Release of Single-Year and Multiyear Estimates

The Census Bureau has released single-year estimates from the full ACS sample beginning with data from the 2005 ACS. ACS 1-year estimates are published annually for geographic areas with populations of 65,000 or more. Beginning in 2008 and encompassing 2005–2007, the Census Bureau will publish annual ACS 3-year estimates for geographic areas with populations of 20,000 or more. Beginning in 2010, the Census Bureau will release ACS 5-year estimates (encompassing 2005–2009) for all geographic areas —down to the tract and block group levels. While eventually all three data series will be available each year, the ACS must collect 5 years of sample before that final set of estimates can be released. This means that in 2008 only 1-year and 3-year estimates are available for use, which means that data are only available for areas with populations of 20,000 and greater.

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

Table 2. **Sets of Sample Cases Used in Producing ACS Multiyear Estimates**

Type of estimate	Year of Data Release				
	2008	2009	2010	2011	2012
	Years of Data Collection				
3-year estimates	2005–2007	2006–2008	2007–2009	2008–2010	2009–2011
5-year estimates	Not Available	Not Available	2005–2009	2006–2010	2007–2011

Source: U.S. Census Bureau.

Differences Between Single-Year and Multiyear ACS Estimates

Currency

Single-year estimates provide more current information about areas that have changing population and/or housing characteristics because they are based on the most current data—data from the past year. In contrast, multiyear estimates provide less current information because they are based on both data from the previous year and data that are 2 and 3 years old. As noted earlier, for many areas with minimal change taking place, using the "less current" sample used to produce the multiyear estimates may not have a substantial influence on the estimates. However, in areas experiencing major changes over a given time period, the multiyear estimates may be quite different from the single-year estimates for any of the individual years. Single-year and multiyear estimates are not expected to be the same because they are based on data from two different time periods. This will be true even if the ACS single year is the midyear of the ACS multiyear period (e.g., 2007 single year, 2006–2008 multiyear).

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

Reliability

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

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

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

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

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

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

Deciding Which ACS Estimate to Use

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

Several key factors can guide users trying to decide whether to use single-year or multiyear ACS estimates for areas where both are available: intended use of the estimates, precision of the estimates, and currency of the estimates. All of these factors, along with an understanding of the differences between single-year and multiyear ACS estimates, should be taken into consideration when deciding which set of estimates to use.

Understanding Characteristics

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

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

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

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

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

For example, consider Sevier County, Tennessee, which had an estimated population of 76,632 in 2004 according to the Population Estimates Program. This population is larger than the Census Bureau's 65,000-population requirement for publishing 1-year estimates. However, many subpopulations within this geographic area will be much smaller than 65,000. Table 5 shows an estimated 21,881 families in Sevier County based on the 2000–2004 multiyear estimate; but only 1,883 families with a female householder, no husband present, with related children under 18 years. Not surprisingly, the 2004 ACS estimate of the poverty rate (38.3 percent) for this subpopulation has a large standard error (SE) of 13.0 percentage points. Using this information we can determine that the CV is 33.9 percent (13.0/38.3).

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

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

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

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

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

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

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

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

Making Comparisons

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

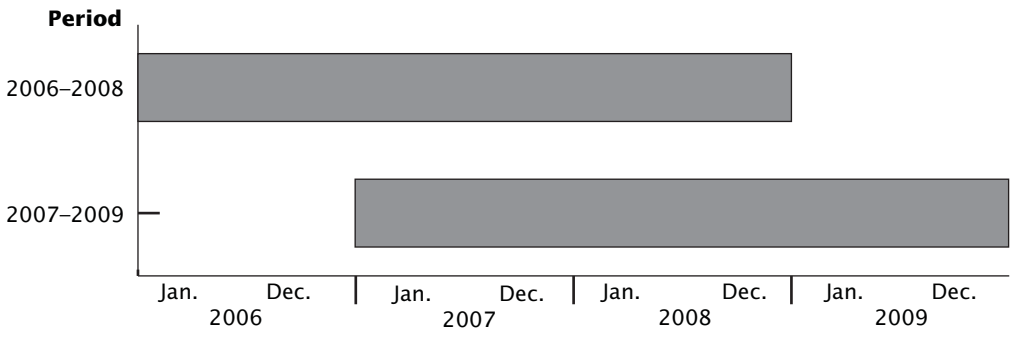
You certainly can compare characteristics for areas with populations of 30,000 to areas with populations of 100,000 but you should use the data set that they have in common. In this example you could use the 3-year or the 5-year estimates because they are available for areas of 30,000 and areas of 100,000.

Assessing Change

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

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

Figure 1. **Data Collection Periods for 3-Year Estimates**

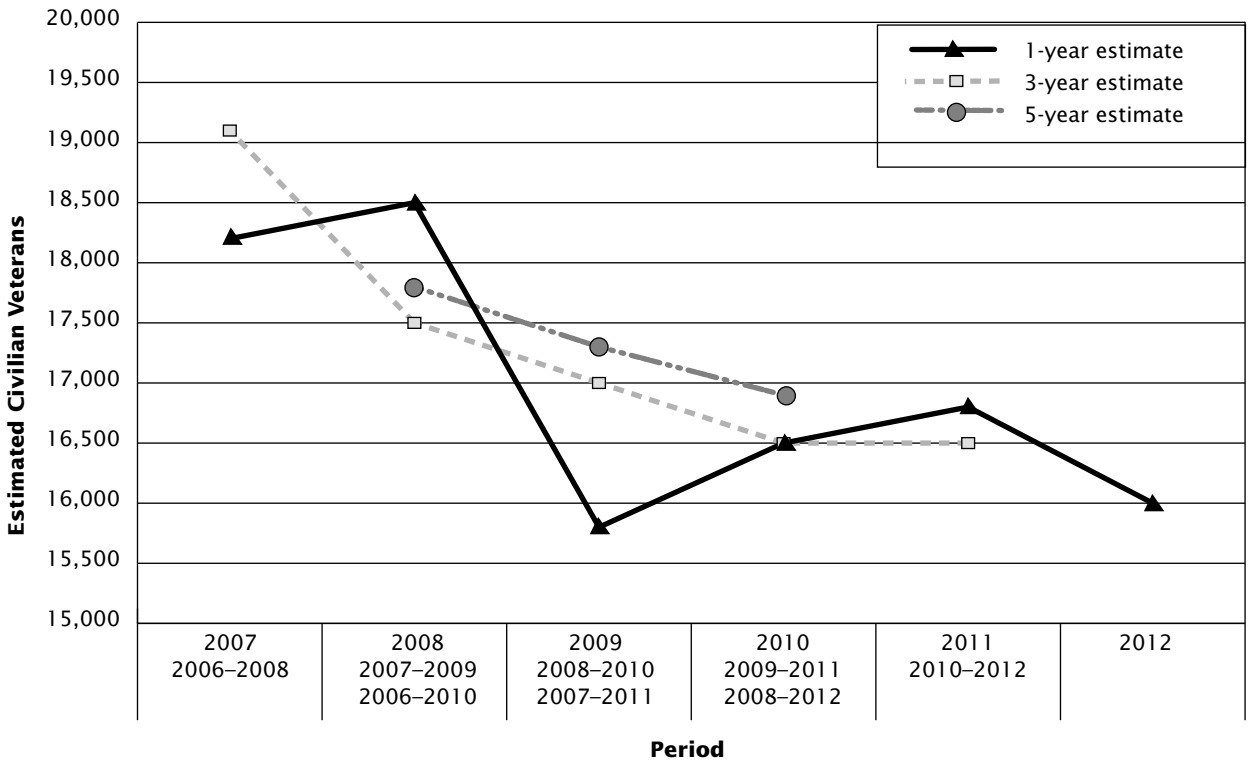


Source: U.S. Census Bureau.

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

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

Figure 2. **Civilian Veterans, County X Single-Year, Multiyear Estimates**



Source: U.S. Census Bureau. Based on data from the Multiyear Estimates Study.

Summary of Guidelines

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

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

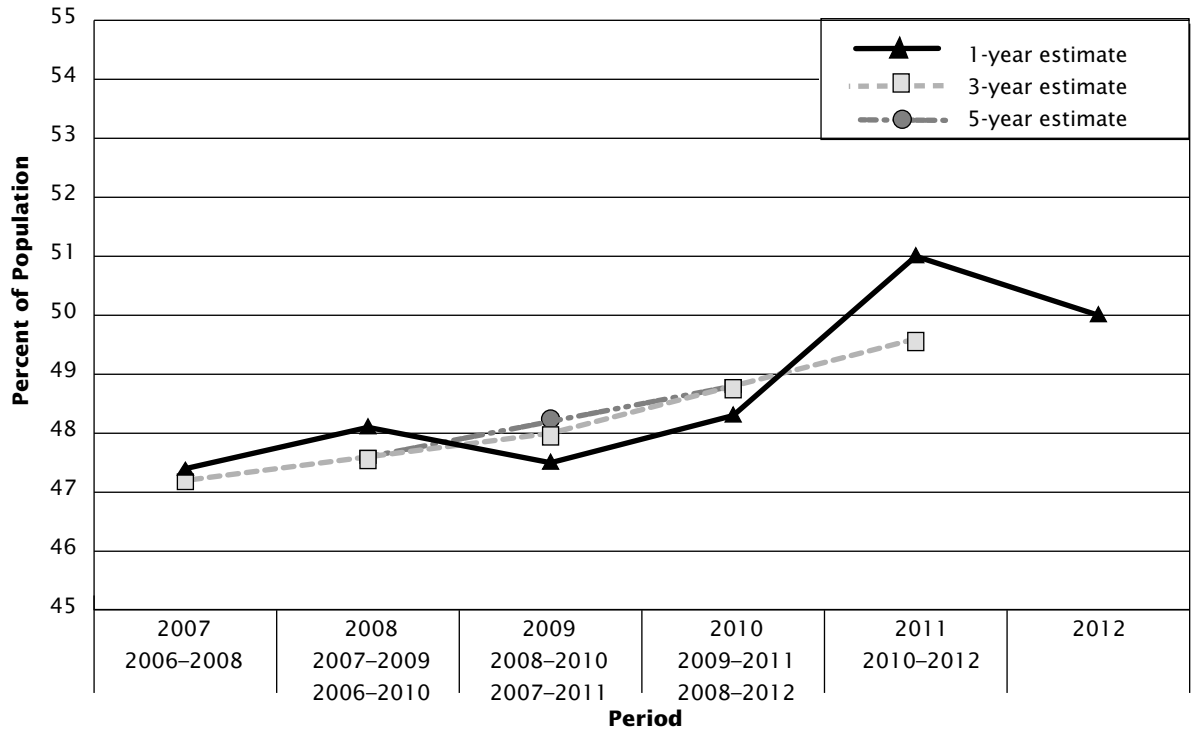
Local area variations may occur due to rapidly occurring changes. As discussed previously, multiyear estimates will tend to be insensitive to such changes when they first occur. Single-year estimates, if associated with sufficiently small CVs, can be very valuable in identifying and studying such phenomena.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

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

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

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

Coefficients of Variation

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

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

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

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

Calculating Coefficients of Variation From Standard Errors

The CV can be expressed as

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

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

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

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

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

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

Deriving Sampling Error Measures From Published MOE

Margin Error (MOE) for Alternate Confidence Levels

$$MOE_{95} = \frac{1.960}{1.645} MOE_{ACS}$$
$$MOE_{99} = \frac{2.576}{1.645} MOE_{ACS}$$

Standard Error (SE)

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

Confidence Interval (CI)

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

Coefficient of Variation (CV)

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

Calculating Margins of Error for Derived Estimates

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

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

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

Calculating MOEs for Aggregated Count Data

To calculate the MOE for aggregated count data:

1) Obtain the MOE of each component estimate.
2) Square the MOE of each component estimate.
3) Sum the squared MOEs.
4) Take the square root of the sum of the squared MOEs.

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

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

where MOE_c is the MOE of the c^{th} component estimate.

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

Table 1. **Data for Example 1**

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

The aggregate estimate is:

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

Obtain MOEs of the component estimates:

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

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

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

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

Calculating MOEs for Derived Proportions

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

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

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

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

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

MOE_{den} is the MOE of the denominator.

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

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

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

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

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

Table 2. **Data for Example 2**

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

The estimated proportion is:

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

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

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

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

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

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

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

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

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

Calculating MOEs for Derived Ratios

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

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

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

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

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

MOE_{den} is the MOE of the denominator.

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

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

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

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

Table 3. Data for Example 3

Characteristic	Estimate	MOE
Black females 25 years and older with a graduate degree (numerator)	4,634	±989
Black males 25 years and older with a graduate degree (denominator)	6,440	±1,328

The estimated ratio is:

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

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

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

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

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

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

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

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

Calculating MOEs for the Product of Two Estimates

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

1) Obtain the MOEs for the two estimates being multiplied together.
2) Square the estimates and their MOEs.
3) Multiply the first squared estimate by the second estimate's squared MOE.
4) Multiply the second squared estimate by the first estimate's squared MOE.
5) Add the results from (3) and (4).
6) Take the square root of (5).

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

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

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

MOE_A is the MOE of the first estimate.

MOE_B is the MOE of the second estimate.

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

Table 4. Data for Example 4

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

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

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

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

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

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

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

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

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

Table 5. **Data for Example 5**

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

The percent change is:

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

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

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

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

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

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

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

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

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

$$1{,}091{,}528{,}529$$

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

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

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

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

Calculating Margins of Error for Derived Estimates

Aggregated Count Data

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

Derived Proportions

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

Derived Ratios

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

Appendix 4.
Making Comparisons

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

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

Determining Statistical Significance

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

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

1) Determine the SE for each estimate (for ACS data, SE is defined by the positive value of the margin of error (MOE) divided by 1.645).[4]
2) Square the resulting SE for each estimate.
3) Sum the squared SEs.
4) Calculate the square root of the sum of the squared SEs.
5) Calculate the difference between the two estimates.
6) Divide (5) by (4).
7) Compare the absolute value of the result of (6) with the critical value for the desired level of confidence (1.645 for 90 percent, 1.960 for 95 percent, 2.576 for 99 percent).
8) If the absolute value of the result of (6) is greater than the critical value, then the difference between the two estimates can be considered statistically significant at the level of confidence corresponding to the critical value used in (7).

Algebraically, the significance test can be expressed as follows:

If $\left|\frac{\hat{X}_1 - \hat{X}_2}{\sqrt{SE_1^2 + SE_2^2}}\right| > Z_{CL}$, then the difference between estimates $\hat{X}_1$ and $\hat{X}_2$ is statistically significant at the specified confidence level, CL

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

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

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

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

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

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

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

[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:

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

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

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

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

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

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

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

Creating Values Used in Multiyear Estimates

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

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

Adjusting Multiyear Estimates Over Time

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

1) Obtain the All Items CPI-U-RS Annual Average for the most current year in each of the time periods being compared.
2) Calculate the inflation adjustment factor as the ratio of the CPI-U-RS Annual Average in (1) from the most recent year to the CPI-U-RS in (1) from the earlier years.
3) Multiply the dollar-denominated estimate for the earlier time period by the inflation adjustment factor.

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

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

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

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

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

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

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

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

Issues Associated With Inflation Adjustment

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

Appendix 6.
Measures of Nonsampling Error

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

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

Types of Nonsampling Errors

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

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

Unit nonresponse is the failure to obtain the minimum required information from a housing unit or a resident of a group quarter in order for it to be considered a completed interview. Unit nonresponse means that no survey data are available for a particular sampled unit or person. For example, if no one in a sampled housing unit is available to be interviewed during the time frame for data collection, unit nonresponse will result.

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

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

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

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

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

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

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

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

ACS Quality Measures

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

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

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

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

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

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

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

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

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

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

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

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

Appendix 7.
Implications of Population Controls on ACS Estimates

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

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

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

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

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

Appendix 8.
Other ACS Resources

Background and Overview Information

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

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

ACS Design, Methodology, Operations

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

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

ACS Quality

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

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

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

Guidance on Data Products and Using the Data

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

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

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

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