

# A Compass for Understanding and Using American Community Survey Data

*What Puerto Rico Community Survey Data Users  
Need to Know*

Issued  
March 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<sup>®</sup>  
**Census**  
**2010**

## Acknowledgments

**Linda A. Jacobsen**, Vice President, Domestic Programs, Population Reference Bureau and **Mark Mather**, Associate Vice President, Domestic Programs, Population Reference Bureau, drafted the *What General Data Users Need to Know* 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., and **Susan P. Love**, Consultant, provided initial review of this handbook.

**Enid Santana** coordinated the adaptation of this handbook for Puerto Rico with assistance from **Francisco Acevedo**, Statistical Project Manager, Puerto Rico Statistical Institute, **Luis Avilés**, President Board of Directors, Puerto Rico Institute of Statistics, **Mario Marazzi**, Executive Director Puerto Rico Institute of Statistics, **Alfredo Navarro**, Assistant Division Chief, Decennial Statistical Studies Division, Census Bureau, and **Idania Rodríguez**, Statistical Project Manager, Puerto Rico Institute of Statistics.

**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**, 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 March 2009

*What Puerto Rico Community Survey Data Users Need to Know*



**U.S. Department of Commerce**  
**Vacant,**  
Secretary

**Vacant,**  
Deputy Secretary

**Economics and Statistics Administration**  
**Vacant,**  
Acting 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 Puerto Rico Community  
Survey Data Users Need to Know*  
U.S. Government Printing Office,  
Washington, DC,  
2009.



### **Economics and Statistics Administration**

**Vacant,**

Acting 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

<b>Foreword</b> .....	iv
<b>Introduction</b> .....	1
<b>About the Puerto Rico Community Survey</b> .....	1
<b>Who Uses the PRCS and Why?</b> .....	2
<b>How the PRCS Works</b> .....	3
<b>Differences Between the PRCS and the Decennial Census</b> .....	3
<b>Geographic Areas Covered in the PRCS</b> .....	4
<b>Benefits and Challenges of PRCS Data</b> .....	7
<b>Understanding Single-Year and Multiyear Estimates</b> .....	7
Making Comparisons With Multiyear Estimates .....	8
<b>Understanding Margin of Error</b> .....	9
<b>Accessing PRCS Data</b> .....	10
<b>Case Studies Using PRCS Data</b> .....	14
Case Study #1: The Local Program Administrator .....	14
Case Study #2: The Consultant.....	16
Case Study #3: The Local Planner.....	19
<b>Avoiding Pitfalls When Working With PRCS Data</b> .....	22
<b>Other Resources for Working With ACS and PRCS Data</b> .....	22
<b>Conclusion</b> .....	23
<b>Glossary</b> .....	24
<b>Appendixes</b> .....	A-1
Appendix 1. Understanding and Using Single-Year and Multiyear Estimates .....	A-1
Appendix 2. Differences Between ACS and Decennial Census Sample Data .....	A-8
Appendix 3. Measures of Sampling Error.....	A-11
Appendix 4. Making Comparisons .....	A-18
Appendix 5. Using Dollar-Denominated Data .....	A-22
Appendix 6. Measures of Nonsampling Error .....	A-24
Appendix 7. Implications of Population Controls on ACS Estimates .....	A-26
Appendix 8. Other ACS Resources .....	A-27

## 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 English and 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](http://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

This handbook provides an overview of the Census Bureau’s Puerto Rico Community Survey (PRCS), which is the equivalent of the American Community Survey for Puerto Rico. The purpose is to help general audiences understand the basics of the PRCS, its opportunities and challenges, and how to access and use the PRCS data on the Census Bureau’s Web site. It

concludes with examples of how PRCS data can be used to answer real questions about demographic and social-economic aspects of the Puerto Rican society. A glossary and a series of technical appendixes are also included at the back of this handbook for those interested in more advanced ACS/PRCS issues and applications.

## About the Puerto Rico Community Survey

The War Department of the United States carried out an enumeration in Puerto Rico in 1899 following the island’s acquisition from Spain in 1898 and there have been decennial censuses from 1910 to the present. Recent censuses have consisted of a “short form,” which included basic questions about age, sex, race, Hispanic origin, household relationship, and owner/renter status, and a “long form” used for only a sample of households that included not only the basic “short-form” questions but also detailed questions about socioeconomic and housing characteristics.

The PRCS is an islandwide survey that started in 2005. It is conducted by the Census Bureau and is designed to provide communities with reliable and timely demo-

graphic, housing, social, and economic data every year. It will replace the decennial census long form in 2010 and thereafter by collecting long-form type information throughout the decade rather than only once every 10 years. The PRCS data will provide a continuous stream of updated information for Puerto Rico and municipio governments and will revolutionize the way we use data to understand our communities.

The PRCS covers a broad range of topics about demographic, housing, social, and economic characteristics of the population as shown in Table 1. As with the decennial census, strict confidentiality laws protect all information that could be used to identify individuals or households.

Table 1. **Subjects Included in the PRCS**

<p><b>Demographic Characteristics</b></p> <ul style="list-style-type: none"> <li>Age</li> <li>Sex</li> <li>Hispanic Origin</li> <li>Race</li> <li>Relationship to Householder (e.g., spouse)</li> </ul> <p><b>Economic Characteristics</b></p> <ul style="list-style-type: none"> <li>Income</li> <li>Food Stamps Benefit</li> <li>Labor Force Status</li> <li>Industry, Occupation, and Class of Worker</li> <li>Place of Work and Journey to Work</li> <li>Work Status Last Year</li> <li>Vehicles Available</li> <li>Health Insurance Coverage*</li> </ul>	<p><b>Social Characteristics</b></p> <ul style="list-style-type: none"> <li>Marital Status and Marital History*</li> <li>Fertility</li> <li>Grandparents as Caregivers</li> <li>Ancestry</li> <li>Place of Birth, Citizenship, and Year of Entry</li> <li>Language Spoken at Home</li> <li>Educational Attainment and School Enrollment</li> <li>Residence One Year Ago</li> <li>Veteran Status, Period of Military Service, and VA Service-Connected Disability Rating*</li> <li>Disability</li> </ul>	<p><b>Housing Characteristics</b></p> <ul style="list-style-type: none"> <li>Year Structure Built</li> <li>Units in Structure</li> <li>Year Moved Into Unit</li> <li>Rooms</li> <li>Bedrooms</li> <li>Kitchen Facilities</li> <li>Plumbing Facilities</li> <li>House Heating Fuel</li> <li>Telephone Service Available</li> <li>Farm Residence</li> </ul> <p><b>Financial Characteristics</b></p> <ul style="list-style-type: none"> <li>Tenure (Owner/Renter)</li> <li>Housing Value</li> <li>Rent</li> <li>Selected Monthly Owner Costs</li> </ul>
---	---	--

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

# Who Uses the PRCS and Why?

The PRCS offers up-to-date information every year about important social and economic issues. The information obtained by the PRCS is used to effectively manage and evaluate the island and local government programs and to monitor year-to-year changes. Puerto Rico and municipio governments are involved in administering and evaluating programs involving federal funding through block grants. Community officials can use the data to monitor the well being of children, families, and the elderly; and to determine where to locate highways and new day care centers. Local businesses also use this vital information to determine if a town has the workforce needed to support business. Nongovernmental organizations also use the PRCS in a variety of ways to monitor changes among important population groups. Journalists use PRCS data to

report on changes in the characteristics of our society or to put a piece of anecdotal evidence into a broader context. The text box provides more examples of how PRCS data are being used.

Much of the PRCS data provided on the Census Bureau's Web site are available separately by age group, race, Hispanic origin, and sex. For example, data users can compare the poverty status of children and the elderly, college enrollment rates for men and women, or housing costs among municipios. No other resource provides such a wealth of social, economic, housing, and demographic information about Puerto Rican society. See Appendix 8 for Web site links to additional background information about the ACS/PRCS.

**State government.** The U.S. Department of Health and Human Services and the Puerto Rico local social services departments use data on age and relationship to plan programs for older people living alone and to develop economic and social programs necessary to support grandparents as caregivers. The Special Supplemental Food Program for Women, Infants, and Children (WIC) uses income data from the PRCS to determine the potential demand for food assistance in Puerto Rico and the municipios.<sup>1</sup>

**Nongovernmental organizations.** The Sociedad de Educación y Rehabilitación (SER) de Puerto Rico uses PRCS data to monitor and provide programs and services to children with disabilities across the island. ASPIRA de Puerto Rico is an organization that uses PRCS data on poverty, educational attainment, school enrollment, and employment status to develop scholarships and grants for leadership programs for children in low-income families.

**Journalists.** An article on income by El Nuevo Día used PRCS data to report that there is a relationship between educational attainment, income level, and workforce participation. The article states that the risk of unemployment decreases when a person stays longer in school and it is reduced even more if they earn an academic degree.<sup>2</sup>

**Community and local planners.** Information from the PRCS is critical for community and local planners, who need up-to-date information about their communities to decide where to locate facilities and services for people with disabilities, evaluate the need for new roads, schools, senior centers, and other basic services. For example, the Puerto Rico Department of Transportation uses information about origins and destinations of journey to work, travel times, and means of transportation to implement the Transportation Equity Act for the 21st Century and other federal and state laws.

<sup>1</sup> For more information about federal uses of ACS/PRCS data, see U.S. Census Bureau, Subjects Planned for the 2010 Census and American Community Survey: Federal Legislative and Program Uses, <[http://www.census.gov/acs/www/Downloads/Final\\_2010\\_Census\\_and\\_American\\_Community\\_Survey\\_Subjects\\_Notebook.pdf](http://www.census.gov/acs/www/Downloads/Final_2010_Census_and_American_Community_Survey_Subjects_Notebook.pdf)>.

<sup>2</sup> El Nuevo Día, "Que nos dice el indicador?," San Juan, Puerto Rico, November 9, 2008, <[http://www.elnuevodia.com/diario/noticia/negocios/negocios/%C2%BFque\\_nos\\_dice\\_el\\_indicador?/479634](http://www.elnuevodia.com/diario/noticia/negocios/negocios/%C2%BFque_nos_dice_el_indicador?/479634)>.



# How the PRCS Works

The PRCS samples about 36,000 addresses each year, resulting in about 22,000 final interviews in Puerto Rico. The annual PRCS sample is smaller than that of the Census 2000 long-form sample. As a result, the PRCS needs to combine population and housing data from multiple years to produce reliable estimates for small municipios and other local areas. To provide information for communities each year, the PRCS will provide 1-year, 3-year, and 5-year estimates.

PRCS data are very timely because they are released in the year immediately following the year in which they are collected. As shown in Table 2, starting with the data collected in 2005, PRCS 1-year estimates have been published for all areas with populations of

65,000 or more. This includes the Commonwealth of Puerto Rico, 12 municipios and 7 cities, among other geographic areas. In December 2008, the first 3-year estimates were released based on data from the 2005, 2006, and 2007 surveys. By December 2009, the PRCS will have sampled approximately 180,000 addresses and, by 2010, the PRCS will provide the first 5-year estimates of demographic, housing, social, and economic data for municipios, cities, and other small geographic areas in Puerto Rico. These 5-year estimates will be updated annually by removing the earliest year and replacing it with the latest one, and, for the first time will provide the ability to monitor social and economic trends in local communities. Spanish versions for these data releases are also available.

Table 2. **Release Schedule for PRCS Data**

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

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

# Differences Between the PRCS and the Decennial Census

While the main function of the decennial census is to provide counts of people for the purpose of legislative redistricting, the primary purpose of the PRCS is to measure the changing social and economic characteristics of the Puerto Rico population by collecting information throughout the decade. As a result, the PRCS does not provide official counts of the population. The Census Bureau's Population Estimates Program will continue to be the official source for Puerto Rico annual population totals, by age and sex. For more information about population estimates for Puerto Rico, visit

the Census Bureau's Web site at <[www.census.gov/popest/puerto\\_rico/](http://www.census.gov/popest/puerto_rico/)> or see Appendix 7 of this handbook. PRCS estimates are controlled at the municipio level to match the Census Bureau's annual population estimates by age and sex.

Although the questions used in the PRCS are very similar to those included on the long form used in Census 2000, there are some important differences between the two surveys. While the decennial census has provided a snapshot of the Puerto Rico population

---

once every 10 years, the PRCS has been described as a “moving video image, continually updated to provide much needed data about” the people “in today’s fast-moving world.”<sup>3</sup>

Because PRCS data are collected continuously, they are not always comparable with data collected from the decennial census. For example, in the case of employment statistics, both surveys ask about employment status during the week prior to the survey. However, data from the decennial census are typically collected between March and June, whereas data from the PRCS are collected nearly every day and reflect employment throughout the year. Differences in these responses may in turn affect data on commuting, occupation, and industry.<sup>4</sup> Other factors that may also have an impact

on the data, such as seasonal variation in population and minor differences in question wording and question order, are described in more detail in Appendix 2 of this handbook. Additional guidance on making comparisons with PRCS and census data is provided in the section on “Making Comparisons” in Appendix 4.

Readers should note that in 2006, the PRCS began including samples of the population living in group quarters (e.g., jails, college dormitories, and nursing homes) for the first time. As a result, 2006 PRCS data may not be comparable with data from 2005 PRCS. This is especially true for estimates of young adults and the elderly, who are more likely than other population groups to be living in group quarters facilities.

## Geographic Areas Covered in the PRCS

---

The PRCS data are tabulated for a variety of geographic areas. Table 3 shows the type and number of geographic areas included in the PRCS, as well as the type of PRCS estimates each will receive. The table is not exhaustive; it only covers major types of geographic areas. The information is based on geographic boundaries and population counts in 2007 and is therefore subject to change in future years.

Since 2006, data were available for geographic areas with at least 65,000 people, including municipios, urban areas, places, metropolitan areas, cities, school districts, and Public Use Microdata Areas.<sup>5</sup> Starting in December 2008, 3-year estimates are available for all areas with at least 20,000 people, and by 2010, 5-year estimates for geographic areas down to the block group level will be available. This means that by 2010, for example, there will be three sets of numbers available for San Juan Municipio—1-year, 3-year, and 5-year estimates. Less populous areas, such as Barrio Miradero in Mayagüez Municipio, will receive only 5-year estimates. As shown in Table 3 the majority of areas will receive only 5-year estimates.

In Figures 1 through 3, sample maps for Puerto Rico show users some of the key geographic areas available through the PRCS: Public Use Microdata Areas, municipios (county equivalent), and tracts.

The Census Bureau divides Puerto Rico into 30 Public Use Microdata Areas, or PUMAs, each of which has a minimum population of 100,000. They are constructed based on municipio, census tract, and place boundaries. Typically, municipios with populations of 200,000 or more are subdivided into multiple PUMAs, while PUMAs in less populous areas encompass groups of contiguous municipios. Because all PUMAs meet the 65,000-population threshold that is needed to produce 1-year PRCS estimates, they provide a useful geographic area for presenting data for less populous areas that otherwise would not receive 1-year estimates. The Census Bureau released PUMA-level data starting in 2006, based on the 2005 PRCS.

Municipios are also important because they are the primary legal subdivision in Puerto Rico. Of Puerto Rico’s 78 municipios (see Figure 2), there are 12 that are large enough to receive 1-year estimates. These areas also receive 3-year and 5-year estimates. Of the remaining 66 municipios, 53 are large enough to receive 3-year estimates. Only 13 out of 78 municipios will only receive 5-year estimates each year starting in 2010.

---

<sup>3</sup> Kathleen B. Cooper. “Halfway to the 2010 Census: The Countdown and Components to a Successful Decennial Census,” paper presented to the House Subcommittee on Federalism and the Census, U.S. House of Representatives, Washington, DC, April 19, 2005, <<http://www.ogc.doc.gov/ogc/legreg/testimon/109f/Cooper0419.htm>>.

<sup>4</sup> U.S. Census Bureau, Meeting 21st Century Demographic Data Needs—Implementing the American Community Survey: Report 5: Comparing Economic Characteristics With Census 2000, Washington DC, 2004, <<http://www.census.gov/acs/www/Downloads/Report05.pdf>>.

<sup>5</sup> Public Use Microdata Areas, or PUMAs, are census-constructed geographic areas, each with approximately a population of 100,000.

Table 3. **Major Geographic Areas and Type of PRCS Estimates Received**

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
Puerto Rico	1	100.0	0.0	0.0
Public Use Microdata Areas	30	100.0	0.0	0.0
Metropolitan statistical areas	8	100.0	0.0	0.0
Micropolitan statistical areas	5	0.0	60.0	40.0
Municipios	78	15.4	67.9	16.7
Urban areas	18	55.6	16.7	27.8
School districts (elementary, secondary, and unified)	1	100.0	0.0	0.0
Places (cities, towns, and census designated places)	225	3.1	4.4	92.4
Census tracts	861	0.0	0.0	100.0
Census block groups	2477	0.0	0.0	100.0
Zip code tabulation areas	134	0.0	0.0	100.0

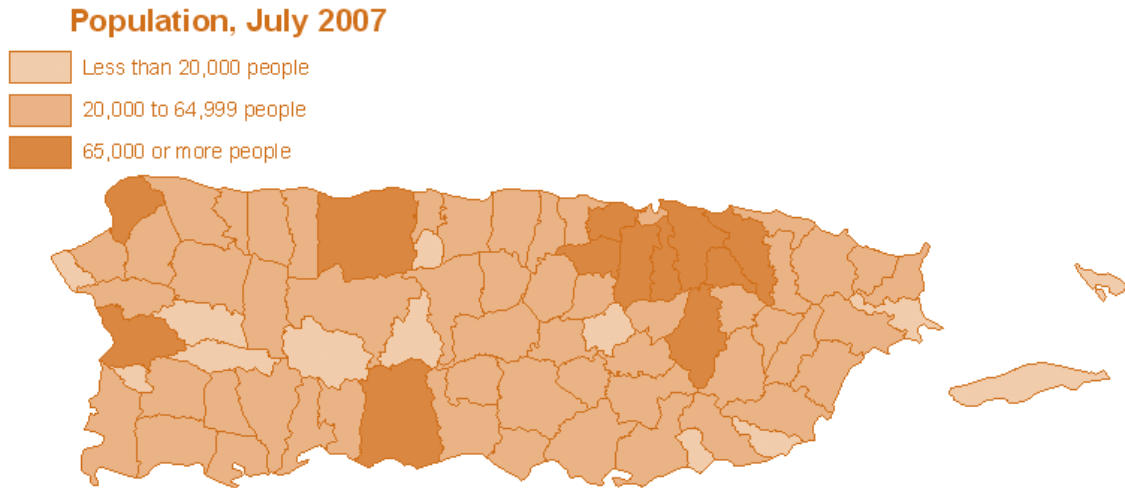
Source: U.S. Census Bureau, 2008. This tabulation is restricted to geographic areas in Puerto Rico. It is based on the population sizes of geographic areas from the July 1, 2007, Census Bureau Population Estimates and geographic boundaries as of January 1, 2007. Because of the potential for changes in 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.

Figure 1. **Census Public Use Microdata Areas in Puerto Rico: 2007**



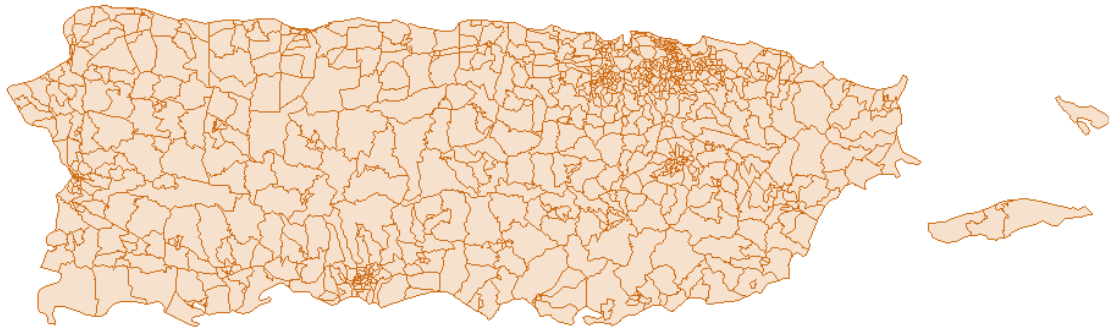
Source: U.S. Census Bureau.

Figure 2. **Municipios in Puerto Rico by Population Size: 2007**



Source: U.S. Census Bureau, 2007 Population Estimates Program.

Figure 3. **Census Tracts in Puerto Rico: 2007**



Source: U.S. Census Bureau.

Census tracts are small geographic areas—with an average of about 4,000 people—that are commonly used to present information for small towns, rural areas, and neighborhoods. PRCS data are not currently

available at the tract level because of their small population sizes. Figure 3 shows the type of geographic detail that will be available once enough data are collected to produce reliable data for census tracts.

# Benefits and Challenges of PRCS Data

---

The PRCS provides several advantages over the information that has been collected in the past through the decennial census long-form samples. The main benefits of the PRCS are timeliness and access to annual data for municipios, local areas, and small population groups. The PRCS will deliver useful, relevant data, similar to data from previous census long forms, but updated every year instead of every 10 years.

Puerto Rico and municipio governments and planners rely on demographic, housing, social, and economic data for their budget formulations and to allocate funds. For example, PRCS data can be used to determine funding levels for nutritional assistance programs or to help decide where to build a day care center. Corporations, small businesses, and individuals can use these data to develop business plans, to set strategies for expansion or starting a business, and to determine if their service areas meet current and future needs.

Small towns and rural communities have much to gain from the PRCS. Lacking the staff and resources to conduct their own research, many local communities rely on decennial census information that becomes increasingly outdated throughout the decade, or use local administrative records that are not comparable with information collected in neighboring areas.<sup>6</sup> The PRCS is the first islandwide survey that can be used to monitor changes in local areas and to make valid, mid-decade comparisons between different communities.

The PRCS also provides tools for those who want to conduct their own research. The PRCS includes a PUMS

file each year that enables researchers to create custom universes and tabulations from individual PRCS records that have been stripped of personally identifiable information. The 2007 PRCS PUMS file includes housing units that were interviewed and included in the published PRCS tabulations for 2007.<sup>7</sup>

The main challenges for PRCS data users are understanding and using multiyear estimates and the relatively large confidence intervals associated with PRCS data for smaller geographic areas and groups of the population. Both of these issues are addressed in more detail below. PRCS data will be produced every year, but in exchange for this benefit, the sample size of the PRCS needs to be smaller than that of the Census 2000 long-form sample.

In late 2010, data users will have access to 5-year estimates of PRCS data. The sample size based on 5-year period estimates of PRCS data is still smaller than the long-form sample in the decennial census, resulting in larger standard errors in the PRCS 5-year estimates. Census Bureau researchers expect that the higher sampling error in the PRCS is offset, to some extent, by a reduction in nonsampling error, due to the use of experienced PRCS interviewers and the refusal to allow survey data to be obtained from neighbors.<sup>8</sup>

Refer to Appendix 8, particularly the section “ACS Design, Methodology, Operations” for more information. Further information on the methodology of the ACS/PRCS is available on the Census Bureau Web site.

## Understanding Single-Year and Multiyear Estimates

---

Multiyear estimates are likely to confuse some data users, in part because of their statistical properties, and in part because this is a new product from the Census Bureau. The PRCS provides municipios and communities that have at least 65,000 residents with 1-year estimates of demographic, housing, social, and economic characteristics—a benefit to government agencies that need to budget and plan for public services like transportation, medical care, and schools. For geographic areas with smaller populations, the PRCS samples too few households to provide reliable

1-year estimates. For these communities, several years of data are pooled together to create reliable 3-year or 5-year estimates.

Note that 1-year, 3-year, and 5-year estimates from the PRCS are all “period” estimates that represent data collected over a period of time (as opposed to “point-in-time” estimates, such as the decennial census, that

---

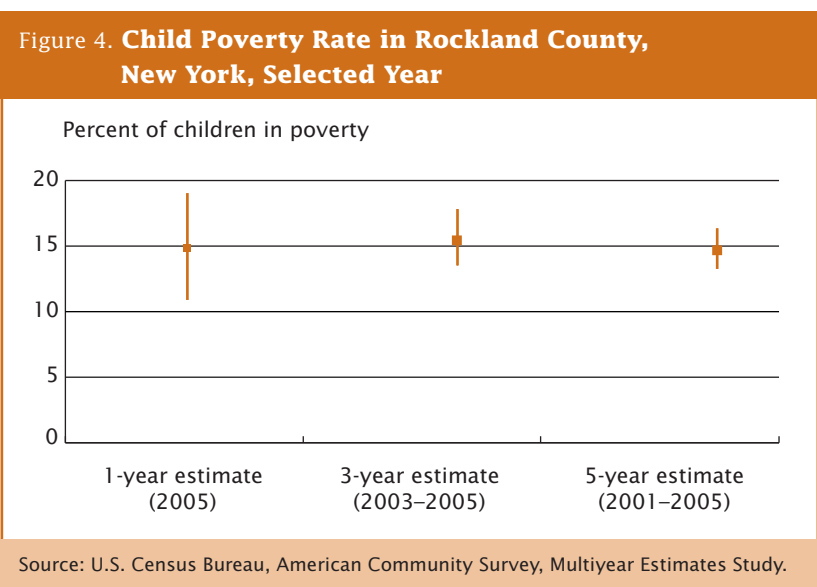
<sup>6</sup> David Swanson and George Hough, “An Evaluation of the American Community Survey: Preliminary Results from a County Level Analysis of the Oregon Test Site,” paper presented at the Annual Meeting of the Mississippi Academy of Sciences, Biloxi, Mississippi, February 19–20, 2004, <[http://www.olemiss.edu/depts/population\\_studies/WorkingPapers/MAS\\_ACS\\_FINAL\\_REV1a.pdf](http://www.olemiss.edu/depts/population_studies/WorkingPapers/MAS_ACS_FINAL_REV1a.pdf)>.

<sup>7</sup> The PRCS includes nearly 36,000 addresses in its initial sample each year. However, a subsample of the addresses that do not respond via mail or telephone is selected to be visited in person, resulting in approximately 23,000 final interviews each year.

<sup>8</sup> Charles H. Alexander, “American Community Survey Data for Economic Analysis,” paper presented to the Census Advisory Committee of the American Economic Association, Washington, DC, October 18–19, 2001, <<http://www.census.gov/acs/www/Downloads/ACS/Paper32.doc>>.

approximate the characteristics of an area on a specific date). While a 1-year estimate includes information collected over a 12-month period, a 3-year estimate represents data collected over a 36-month period, and a 5-year estimate includes data collected over a 60-month period. Therefore, PRCS estimates based on data collected from 2005–2007 should not be called “2006” or “2007” estimates. Nor should 2005–2009 period estimates be labeled “2007” estimates, even though that is the midpoint of the 5-year period. Multiyear estimates should be labeled to indicate clearly the full period of time (e.g., “The child poverty rate in 2005–2007 was X percent”). For more information about understanding single-year and multiyear estimates, see Appendix 1 of this handbook.

The primary advantage of using multiyear estimates is the increased statistical reliability of the data compared with that of single-year estimates, particularly for smaller geographic areas and small population groups. A good illustration comes from the Multiyear Estimates Study—a research project undertaken by the Census Bureau to study the characteristics of multiyear estimates. Although this study did not include any geographic areas in Puerto Rico, we can draw conclusions about PRCS data from test areas in the United States. An example of 1-, 3-, and 5-year estimates for Rockland County, New York, is shown in the chart in Figure 4. The lines above and below the point estimates represent the confidence intervals, or ranges of uncertainty, around each estimate. As you can see, the level of precision improves dramatically with the 3- and 5-year estimates as indicated by the shorter lines.



### Making Comparisons With Multiyear Estimates

One of the challenges for data users is deciding how to compare geographic areas with different population sizes. Estimates for areas with fewer than 20,000 people are produced only in the form of 5-year estimates, while 1-year estimates are available for areas with populations of 65,000 or more. To facilitate such geographic comparisons, the Census Bureau plans to publish 5-year estimates for all geographic areas and 3-year estimates for all areas with 20,000 or more residents. The Census Bureau recommends comparing estimates based only on the same type of estimates. For example, suppose you want to compare estimates for Barrio Playa Guayanés in Yabucoa Municipio, a small place on the east coast of the island, with estimates for Caguas Municipio. Even though the PRCS publishes 1-year estimates for Caguas, only 5-year estimates will be published for Barrio Playa Guayanés.

Thus, in 2010, when 5-year estimates for smaller geographic areas become available, you should compare 2005–2009 estimates for Barrio Playa Guayanés with 2005–2009 estimates for Caguas Municipio, even though more recent 1-year estimates will be available for Caguas.

While multiyear estimates provide valuable data for small geographic areas, using them to look at trends for small populations can be challenging because they rely on pooled data for 3 or 5 years. For example, comparisons of 3-year estimates from 2005–2007 and 2006–2008 are unlikely to show much difference because two of the years overlap (both sets of estimates include the same data collected in 2006 and 2007). The overlap is even greater for the 5-year estimates used for areas with fewer than 20,000 people.

The Census Bureau suggests comparing periods that do not overlap—comparing 2005–2007 estimates with 2008–2010 estimates, for example. This means waiting longer to identify a trend. However, in areas undergoing fundamental shifts in the size or composition of the population, change may be so substantial that it will be obvious after only a few years. Additional guidance on the use and interpretation of single-year and multiyear estimates is provided in Appendix 1.

Data users also need to use caution in looking at trends involving income or other measures that are adjusted for inflation, such as income and monetary variables. Appendix 5 provides information on the adjustment of single-year and multiyear ACS/PRCS estimates for inflation. Note that inflation adjustment is based on a national-level consumer price index: it does not adjust for differences in costs of living across different geographic areas.

## Understanding Margin of Error

All data that are based on samples, such as the PRCS and the census long-form samples include a range of uncertainty. Two broad types of error can occur: sampling error and nonsampling error. Nonsampling errors can result from mistakes in how the data are reported or coded, problems in the sampling frame or survey questionnaires, or problems related to nonresponse or interviewer bias. The Census Bureau tries to minimize nonsampling errors by using trained interviewers and by carefully reviewing the survey's sampling methods, data processing techniques, and questionnaire design. Appendix 6 includes a more detailed description of different types of errors in the ACS/PRCS and other measures of ACS/PRCS quality.

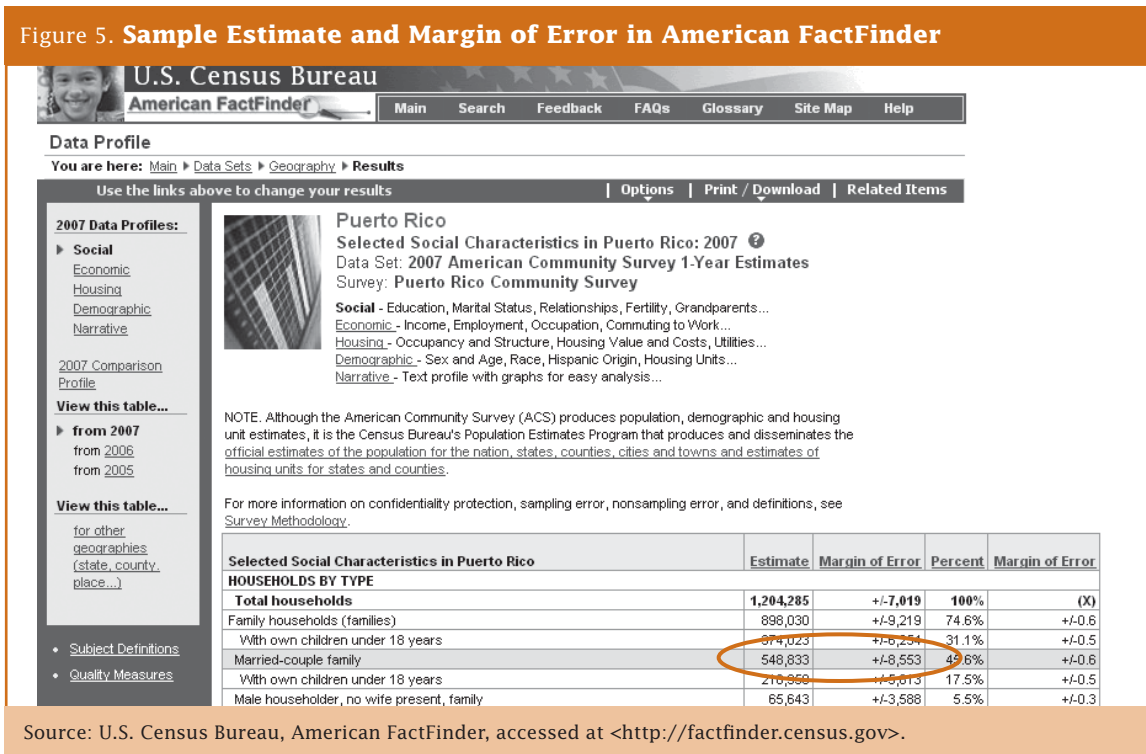
Sampling error occurs when data are based on a sample rather than the full population. Sampling error is easier to measure than nonsampling error and can be used to assess the statistical reliability of survey

data. For any given area, the larger the sample, the greater the confidence in the estimate. The Census Bureau reported the 90-percent confidence interval on all PRCS estimates produced for 2005. The 90-percent confidence intervals define a range expected to contain the true value of an estimate with a level of confidence of 90 percent. Margins of error are easily converted into these confidence ranges. With the release of the 2006 PRCS data, margins of error are now provided for every PRCS estimate. For example, the 2007 PRCS data profile for Puerto Rico, shown in Figure 5, shows that 548,833 married-couple families resided in Puerto Rico in 2007. By adding and subtracting the margin of error from the point estimate, we can calculate the 90-percent confidence interval for that estimate:

$$548,833 - 8,553 = 540,280 = \text{Lower-bound interval}$$

$$548,833 + 8,553 = 557,386 = \text{Upper-bound interval}$$

Figure 5. Sample Estimate and Margin of Error in American FactFinder



Therefore, we can be 90-percent confident that the true number of married-couple families in Puerto Rico falls somewhere between 540,280 and 557,386. Detailed information about sampling error and instructions for calculating confidence intervals and margins of error are included in Appendix 3 of this handbook.

The margin of error around an estimate is important because it helps you draw conclusions about the data. Small differences between two estimates may not be statistically significant if the confidence intervals of those estimates overlap. However, the Census Bureau cautions data users not to rely on overlapping confidence intervals as a test for statistical significance, because this method will not always produce accurate results. Instead, the Census Bureau recommends following the detailed instructions provided in Appendix 4 of this handbook for conducting statistical significance tests.

In some cases, data users will need to construct custom PRCS estimates by combining data across multiple

geographic areas or population groups or it may be necessary to derive a new percentage, proportion, or ratio from published PRCS data. In such cases, additional calculations are needed to produce confidence intervals and margins of error for the derived estimates. Appendix 3 provides detailed instructions on how to make these calculations. Note that these error measures do not tell us about the magnitude of non-sampling errors.

Some advanced data users will also want to construct custom PRCS estimates from the PRCS' Public Use Microdata Sample (PUMS). There are separate instructions for conducting significance tests for PUMS estimates, available on the Census Bureau's American FactFinder Web site at <[http://factfinder.census.gov/home/en/acs\\_pums\\_2007.html](http://factfinder.census.gov/home/en/acs_pums_2007.html)>. (Click on 2007 PUMS Accuracy if using the 2007 ACS PUMS.) Readers can also consult the Public Use Microdata Sample handbook in this handbook series.

## Accessing PRCS Data

All PRCS data are available through the Census Bureau's American FactFinder Web site at <<http://factfinder.census.gov>>.<sup>9</sup> The link to the Puerto Rico Community Survey data sets can be found under the

*Getting Detailed Data* section on the American FactFinder home page. For the Spanish language version of the PRCS click on *en español*, as shown in Figure 6.<sup>10</sup>

Figure 6. American FactFinder Web Site



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

<sup>9</sup> The only exceptions are the ACS/PRCS Summary Tables available for download through the Census Bureau's FTP server, at <[http://www2.census.gov/acs2006/Final\\_Summaryfile/](http://www2.census.gov/acs2006/Final_Summaryfile/)>. These files are designed for advanced users with data processing and or programming capabilities.

<sup>10</sup> Data for Puerto Rico can be accessed from both the ACS and the Puerto Rico Community Survey data sets.



Figure 7 is a screenshot of the data sets page and the links to the various data products available from the 2007 PRCS. For the 2007 PRCS, two data sets are shown—the 3-year estimates (based on the 2005–2007 PRCS) and the 1-year estimates (based on the 2007 PRCS). Tabs also index 2006 and 2005 PRCS data sets. The AFF will default to the most recent data set but users can choose from any of these data sets. For 2007 data, AFF will default to the 3-year estimates data set. It is important for all users to understand that once a data set is selected, the accessed tables will

all correspond to this specific data set. All tables are clearly labeled, identifying the data set.

As shown in Figure 7, a link to explain 1-year vs. 3-year estimates is available. This link offers basic guidance to users about which estimates to select with a basic comparison of 1-year and 3-year estimates. Also available is the 2007 Quick Guide, which provides detailed descriptions of the different ACS/PRCS data products that are available.

**Figure 7. Data Sets Page**

U.S. Census Bureau  
American FactFinder

Main Search Feedback FAQs Glossary Site Map Help

Main ► Data Sets

POPULATION FINDER  
FACT SHEET  
PEOPLE  
HOUSING  
BUSINESS AND GOVERNMENT  
ABOUT THE DATA  
DATA SETS  
• Decennial Census  
• American Community Survey  
• Puerto Rico Community Survey  
• Annual Population Estimates  
• Economic Census  
• Annual Economic Surveys  
DOWNLOAD CENTER  
MAPS  
TOOLS AND REFERENCES

AMERICAN COMMUNITY SURVEY [Load Query](#) | [Clear all selections](#)

The American Community Survey is a nationwide survey designed to provide communities a fresh look at how they are changing. The Puerto Rico Community Survey is the equivalent of the American Community Survey for Puerto Rico. [more...](#)

[2007 Quick Guide](#) | [Errata Notes](#)

**Other Resources**

- [American Community Survey Main Page](#)
- [Quality Measures](#)
- [Public Use Microdata Sample \(PUMS\) - download data and view documentation](#)
- [Download Center](#)
- [Download 1996-2004 data via FTP](#)

[Explain Table and Map Formats](#)

2007

**2005-2007 American Community Survey 3-Year Estimates**

- Data from the American Community Survey and the Puerto Rico Community Survey
- Collected during calendar years 2005, 2006 and 2007
- Available for geographic areas with populations of **20,000 or more**

[Explain 1-year vs. 3-year estimates](#)

**2007 American Community Survey 1-Year Estimates**

- Data from the American Community Survey and the Puerto Rico Community Survey
- Collected during calendar year 2007
- Available for geographic areas with populations of **65,000 or more**

Select from the following:

- [Data Profiles](#)
- [Selected Population Profiles](#)
- [Subject Tables](#)
- [Detailed Tables](#)
- [Geographic Comparison Tables](#)
- [Thematic Maps](#)
- [Reference Maps](#)
- [Custom Table](#)
- [Enter a table number](#)
- [List all tables](#)
- [List all maps](#)
- [Download PUMS data](#)
- [About this data set](#)

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

---

Many of the PRCS data products are similar to those available for Census 2000, but several new data products are also available. All of the tables and graphs are easy to use and include measures of statistical reliability. The set of PRCS/ACS data products are described below. Table 4 provides a quick summary of key PRCS data products.

- **Data profiles and narrative profiles.** The data profiles provide a good starting place for novice data users. Data profiles provide separate fact sheets on the social, economic, demographic, and housing characteristics for different geographic areas. Narrative profiles also provide clear, concise textual descriptions of the data included in the data profiles. They must be accessed by selecting data profiles from the menu.
- **Geographic comparison tables.** Those interested in geographic comparisons for areas within Puerto Rico may be interested in the geographic comparison tables, which allow comparison of PRCS data across a variety of geographic areas, including metropolitan areas, places, and municipios.
- **Thematic maps.** The thematic maps provide graphic displays of the data available through the geographic comparison tables. Different shades of color are used to display variations in the data across geographic areas. Data users can also highlight areas with statistically different values from a selected municipio or metropolitan area of interest.
- **Subject tables.** For information about a particular topic (e.g., employment, education, and income), start with the subject tables, which provide pretabulated numbers and percentages for a wide variety of topics, often available separately by age and sex.
- **Selected population profiles.** Detailed race/ethnic data are available through the selected population profiles. Beginning with the 2007 PRCS, the selected population profiles also included country of birth. Due to population thresholds, most of these PRCS data tables will be available with the release of the 5-year estimates.
- **Comparison profiles.** The comparison profiles show data side-by-side from the 2006 PRCS and the 2007 PRCS, indicating where there is a statistically significant difference between the two sets of estimates. Comparison profiles are only available for 1-year estimates.
- **Detailed tables and summary files.** The detailed tables are the best source for advanced data users or those who want access to the most comprehensive PRCS tables. For more advanced users, detailed tables are also available for download through the ACS Summary File on the Census Bureau's FTP Web site <[www.census.gov/acs/www/Special/acsftp.html](http://www.census.gov/acs/www/Special/acsftp.html)>.
- **Public Use Microdata Samples.** Those with expertise in using statistical analysis software packages may also be interested in the Public Use Microdata Sample (PUMS) files, which contain a sample of individual records of people and households that responded to the survey (stripped of all personal identifiable information). The PUMS files permit analysis of specific population groups and custom variables that are not available through the American FactFinder. For example, PUMS data users can look at the proportion of children ages 5 to 11 living in low-income working families or the number of chemical engineers making \$100,000 in San Juan. More information about the PUMS is available in the PUMS Handbook.
- **Custom tables.** The custom tables option allows you to customize specific detailed tables to meet your needs. You can extract selected rows of data from one or more detailed tables to create a table with just the estimates you want to include.

Table 4. **Summary of PRCS Data Products**

<b>Data product</b>	<b>Geographic areas covered</b>	<b>Description</b>
Data profiles	All*	Provide broad social, economic, housing, and demographic profiles.
Narrative profiles	All*	Summarize the information in the data profiles using concise, nontechnical text.
Selected population profiles	All*	Provide social, economic, and housing profiles for a number of race, ethnic, ancestry, and country of birth groups.
Ranking tables	States, DC, Puerto Rico	Provide state rankings of estimates across key variables.
Subject tables	All*	Similar to data profiles but include more detailed PRCS data, classified by subject.
Detailed tables	All*	Provide access to the most detailed PRCS data and cross-tabulations of PRCS variables.
Geographic comparison tables	All*	Compare geographic areas other than states (e.g., municipios) for key variables.
Thematic maps	All*	Interactive, online maps that can be used to display PRCS data.
Custom tables	All*	Allow the user to extract specific rows of data from the PRCS detailed tables.
Summary files	All*	Provide access to the detailed tables through a series of comma-separated text files on the Census Bureau's FTP site <www2.census.gov>.
Public Use Microdata Sample files	Puerto Rico, PUMAs	Provide access to PRCS microdata for data users with statistical analysis experience.

\*Note: Data will be available for local areas and small population subgroups with the release of 3-year and 5-year estimates.

# Case Studies Using PRCS Data

By now you should have a fairly good understanding of the PRCS, the opportunities it provides, and its limitations. In this last section, we provide some concrete

examples of how PRCS data can be used to answer specific questions about social and economic patterns in Puerto Rico.

## Case Study #1: The Local Program Administrator

Jaime is the administrator of a local program in the Ponce Municipio. In 2006, the city put into action a plan to reduce unemployment rates through a series of seminars and job training initiatives that help the unemployed develop skills that correspond with available job opportunities. He wants to know if these initiatives have reduced the level of unemployment in the city.

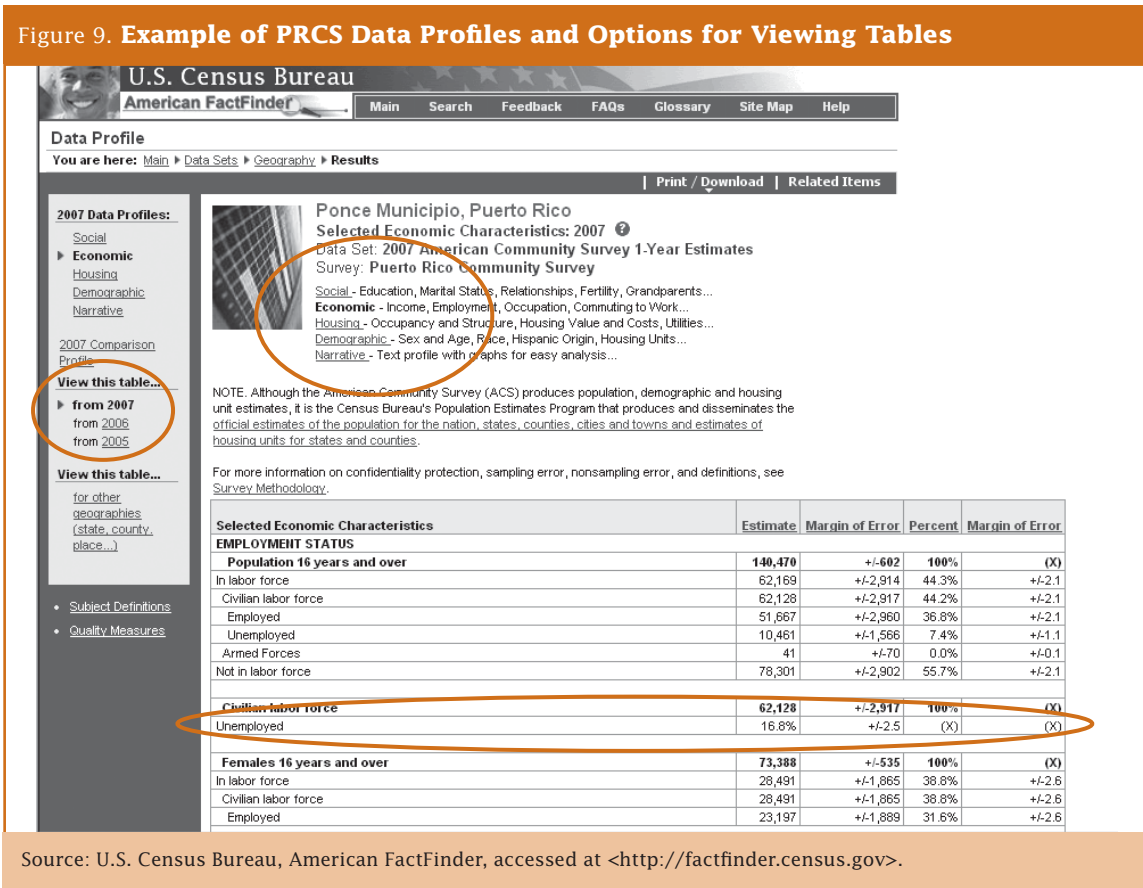
Jaime starts his analysis by going to the American FactFinder and selecting the 2006 PRCS Data Profile for Ponce Municipio (see Figure 8).

The 2006 PRCS Data Profile has five summaries (see circled list at top of Figure 9), including one that focuses on economic characteristics. The economic profile indicates that Ponce's unemployed rate was 18.8 percent with a margin of error of  $\pm 3.0$  percentage points. Next, he uses the navigation tools in the left sidebar to view the same PRCS table for the 2007 PRCS (see Figure 9). Jaime finds that the unemployment rate was 16.8 percent with a margin of error of 2.5 percentage points.

Figure 8. PRCS Data Profiles Available Through American FactFinder

The screenshot shows the American FactFinder website interface. At the top, there is a navigation bar with links for Main, Search, Feedback, FAQs, Glossary, Site Map, and Help. A left sidebar contains a 'Data Sets' menu with categories like POPULATION FINDER, FACT SHEET, PEOPLE, HOUSING, BUSINESS AND GOVERNMENT, ABOUT THE DATA, DATA SETS, DOWNLOAD CENTER, MAPS, and TOOLS AND REFERENCES. The main content area is titled 'AMERICAN COMMUNITY SURVEY' and includes a description of the survey. Below the description, there are two data sets listed: '2006 American Community Survey' and '2005 American Community Survey'. The 2006 entry is circled in orange. To the right of the 2006 entry, there is a 'Select from the following:' section with a list of options, including 'Data Profiles', 'Selected Population Profiles', 'Ranking Tables', 'Subject Tables', 'Detailed Tables', 'Geographic Comparison Tables', 'Thematic Maps', 'Reference Maps', 'Custom Table', 'Enter a table number', 'List all tables', 'List all maps', 'Download PUMS data', and 'About this data set'. The 'Data Profiles' and 'Selected Population Profiles' options are also circled in orange. At the bottom of the page, there is a source citation: 'Source: U.S. Census Bureau, American FactFinder, accessed at <http://factfinder.census.gov>.'

Figure 9. Example of PRCS Data Profiles and Options for Viewing Tables



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

He pastes the poverty data from these tables into a spreadsheet, including the confidence intervals around each estimate.<sup>11</sup> Based on the point estimates, it appears that the unemployment rate has declined in Ponce since the initiative started.

Jaime is interested in the overall change from 2006, but he is not sure from the data whether the decrease over this time period represents a statistically significant change. He uses the instructions included in Appendix 4 of this handbook to find the formulas he needs to test for a significant difference between two PRCS estimates. First, he needs to calculate the standard error associated with each of the two estimates:

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

Where  $MOE_{PRCS}$  is the PRCS published margin of error for the estimate.

For the 2007 estimate:  $SE = \frac{2.5}{1.645} = 1.52$

For the 2006 estimate:  $SE = \frac{3.0}{1.645} = 1.82$

<sup>11</sup> One way to present confidence intervals in Excel is to use "Stock" charts to display the data.

He then uses a second formula to test whether the difference between the two estimates is significant at the 90-percent confidence level.

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

Where  $X_1$  is the 2007 estimate and  $X_2$  is the 2006 estimate,  $SE_1$  is the standard error for the 2007 estimate and  $SE_2$  is the standard error for the 2006 estimate.

$Z_{CL}$  is the critical value for the desired confidence level (=1.645 for 90 percent).

$$\left| \frac{16.8 - 18.8}{\sqrt{(1.52)^2 + (1.82)^2}} \right| > 1.645$$

Since 0.84 is smaller than 1.645, the difference between the two estimates is not statistically significant and Jaime cannot conclude that the unemployment rate in Ponce Municipio has dropped since 2006. He tells his constituents that the data show no change in the unemployment rate but he will continue to monitor these data to see if the 2008 PRCS suggests that unemployment has dropped.

## Case Study #2: The Consultant

Antonio is an independent statistical consultant for real estate investors. He received a call from Mario, one of his best clients, asking for up-to-date information about urban areas with the highest median home value in Puerto Rico. Antonio explains to his client the current limitations of the PRCS data. Right now, he can only get information for geographic areas with at least 20,000 people, and there are margin of errors associated with all of the PRCS estimates, particularly for smaller geographic areas. But he tells Mario he will see what he can find.

Antonio starts by going to the American FactFinder and pulling up the detailed tables from the 2005–2007 ACS 3-year estimates data set, as shown in Figure 10.

The default menu allows him to select geographic areas within a specific state, but the geo within geo menu provides additional options for selecting geographic areas. For example, Antonio can select all municipios within Puerto Rico or output data for places within metropolitan statistical areas, urban areas, or municipios.

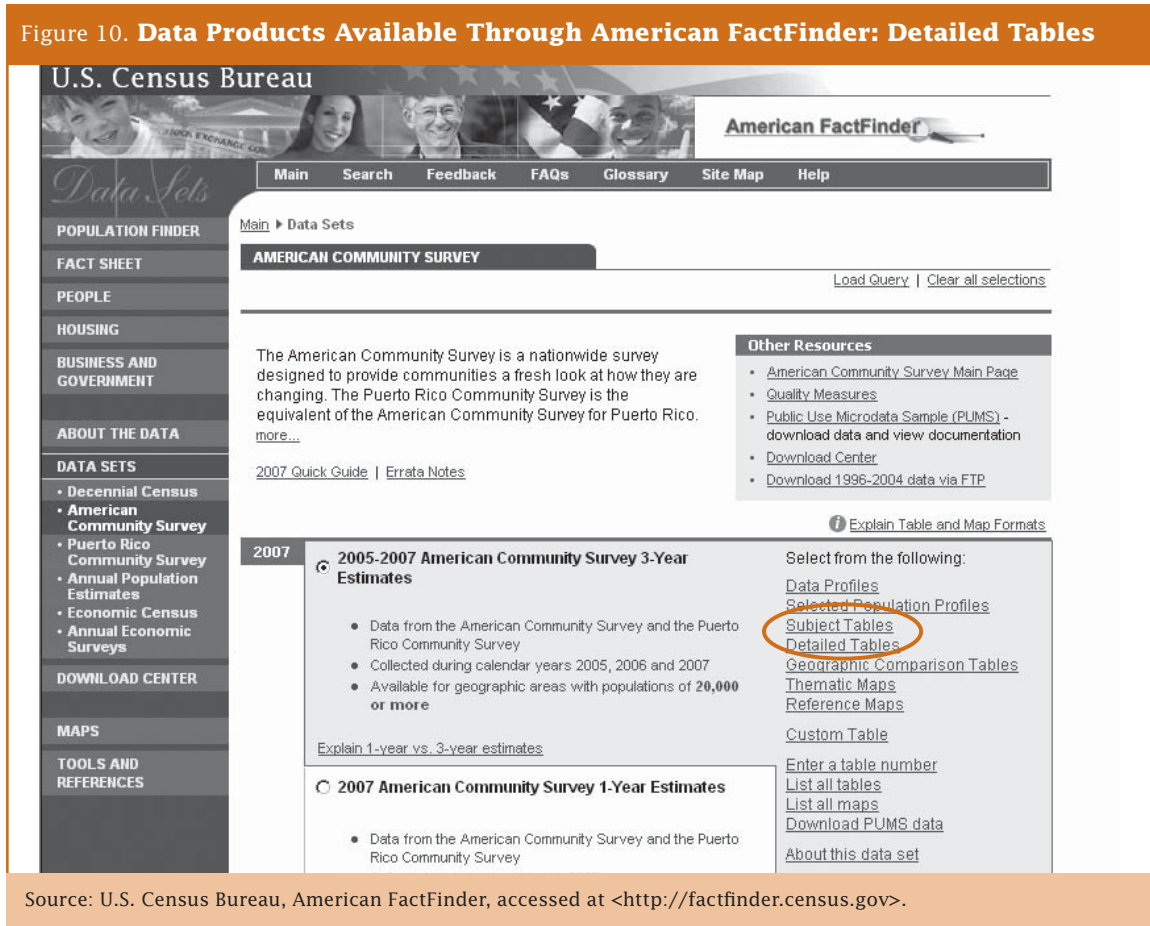


Figure 11 shows the geo within geo tab, which allows him to extract data for all large places (this includes cities and towns with a population of 20,000 or more) in Puerto Rico. He uses the drop-down menu to select Places, Puerto Rico, and All Places. On the next page, he browses the *by subject* menu and selects the Value of Home (Owned and Vacant) topic, which is listed under Housing Characteristics. He selects Table B25077 (Median Value (Dollars) for Owner-Occupied Housing Units) and clicks on the button to show the results. However, since data are displayed online only for the first few places, he decides to use the *Print/*

*Download tab* to export the full table into a spreadsheet (see Figures 12 and 13).

Given Mario's criteria, Guaynabo urban area topped the list. Rather than just considering data for that single place, Antonio puts together the following table (Table 5) showing Median Home Values for the available urban areas in Puerto Rico. He also warns Mario about the margin of error around the estimates. Finally, Antonio directs Mario to the data profiles on the American FactFinder to learn more about the characteristics of people and housing in these selected areas.

Figure 11. Using the “geo within geo” Tab in American FactFinder

The screenshot shows the 'Select Geography' interface. At the top, there's a navigation bar with 'Main', 'Search', 'Feedback', 'FAQs', 'Glossary', 'Site Map', and 'Help'. Below that, the breadcrumb trail reads: 'You are here: Main > Data Sets > Data Sets with Detailed Tables > Geography > Tables > Results'. The main heading is '2005-2007 American Community Survey 3-Year Estimates, Detailed Tables'. Under 'Choose a selection method', the 'geo within geo' tab is circled in orange. Below this, there are instructions to 'Select a geographic type' (set to 'Place'), 'Select a state' (set to 'Puerto Rico'), and 'Select one or more geographic areas and click 'Add''. A list of places is shown, with 'Map It' and 'Add' buttons. At the bottom, 'Current geography selections' are listed, and there are 'Remove' and 'Next' buttons.

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

Figure 12. Print/Download Data From American FactFinder

The screenshot shows the 'Detailed Tables' page. The breadcrumb trail is 'You are here: Main > Data Sets > Data Sets with Detailed Tables > Geography > Tables > Results'. Below this, there are links for 'Options', 'Print / Download' (circled in orange), and 'Related Items'. A note says: 'Note: use download to retrieve all selected tables and geographies'. The data set is identified as 'B25077: MEDIAN VALUE (DOLLARS) - Universe: OWNER-OCCUPIED HOUSING UNITS' for the '2005-2007 American Community Survey 3-Year Estimates' in 'Puerto Rico Community Survey'. It indicates 'geographies 1.5 of 16' with a 'Next' button. A note explains that the ACS produces population, demographic, and housing unit estimates. A link to 'Survey Methodology' is provided. At the bottom, a table shows median values for five areas: Arecibo, Bayamón, Caguas, Carolina, and Cataño.

	Arecibo zona urbana, Puerto Rico		Bayamón zona urbana, Puerto Rico		Caguas zona urbana, Puerto Rico		Carolina zona urbana, Puerto Rico		Cataño zona urbana, Puerto Rico	
	Estimate	Margin of Error	Estimate	Margin of Error	Estimate	Margin of Error	Estimate	Margin of Error	Estimate	Margin of Error
Median value (dollars)	92,000	+/-2,326	131,000	+/-1,897	119,100	+/-1,811	142,600	+/-2,065	107,000	+/-6,362

Source: U.S. Census Bureau, 2005-2007 American Community Survey

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

Figure 13. Downloading Data From American FactFinder Into MS Excel

**Detailed Tables ▸ Download**

■ Select download options and click 'OK' [i Explain my choices](#)

Download file format (data rows plus headings and footnotes)

- Comma delimited (.csv)  
(.csv is compatible with spreadsheet programs such as Microsoft Excel)
- Tab delimited (.lst)
- Rich Text Format (.rtf)
- Microsoft Excel (.xls)

Options (only applies to presentation formats)

- Only the tables and geographies on the screen

Database compatible (data rows only) - the download file is a zip file containing one or more data files and a geographic content file. Geographic codes are included in the database compatible download.

- Microsoft Excel (.xls)
- Comma delimited (.txt)
- Pipe delimited (.txt)

Options (only applies to database compatible formats)

- Include descriptive data element names

[i Using FactFinder with Windows XP Service Pack 2](#)

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

Table 5. Median Home Value in Dollars for Places in Puerto Rico With Population 20,000 or More People: 2005–2007

	Median home value, 2005–2007	Margin of error (+/–)
Guaynabo zona urbana, Puerto Rico	\$210,900	\$6,238
Trujillo Alto zona urbana, Puerto Rico	\$157,700	\$5,030
San Juan zona urbana, Puerto Rico	\$153,800	\$2,286
Carolina zona urbana, Puerto Rico	\$142,600	\$2,065
Levittown comunidad, Puerto Rico	\$136,800	\$3,675
Bayamón zona urbana, Puerto Rico	\$131,000	\$1,897
Caguas zona urbana, Puerto Rico	\$119,100	\$1,611
Cataño zona urbana, Puerto Rico	\$107,000	\$6,362
Humacao zona urbana, Puerto Rico	\$101,900	\$6,140
Fajardo zona urbana, Puerto Rico	\$95,500	\$1,958
Vega Baja zona urbana, Puerto Rico	\$94,200	\$2,828
Ponce zona urbana, Puerto Rico	\$93,400	\$1,445
Arecibo zona urbana, Puerto Rico	\$92,000	\$2,326
Guayama zona urbana, Puerto Rico	\$91,400	\$3,627
Yauco zona urbana, Puerto Rico	\$88,400	\$3,502
Mayagüez zona urbana, Puerto Rico	\$86,000	\$2,356

Note: Because of sampling variability, some of the estimates in this table may not be statistically different from one another or from estimates for other geographic areas not listed in the table.

Source: U.S. Census Bureau, 2005–2007 American Community Survey.



### Case Study #3: The Local Planner

As a member of the planning commission of the Department of Transportation of the Commonwealth of Puerto Rico, Virginia is interested in the changing characteristics of workers and the means of transportation of workers who work in the San Juan Metropolitan Area. She wants to know how workers who drive to work alone compare with those who take public transportation and how this might affect the demand for transportation services in the coming years. In particular, she wants to compare the proportion of workers 16 years and over for whom poverty status is determined and who drive to work alone with those who take public transportation under the same circumstances.

Virginia starts by going to the 2007 Detailed Tables in the American FactFinder. First she selects Metropolitan Statistical Areas/Micropolitan Statistical Areas and then San Juan-Caguas-Guaynabo PR Metro Area from the list of geographic areas available (see Figure 14).

On the next page Virginia uses the *by keyword* tab to pull up a list of all of the transportation tables (see Figure 15). From that list, she chooses Table B08122 (Means of Transportation to Work by Poverty Status in the Past 12 Months). The next page shows her a table with the results.

**Figure 14. Selecting Geographies in American FactFinder**

U.S. Census Bureau  
American FactFinder

Main Search Feedback FAQs Glossary Site Map Help

Select Geography

You are here: Main > Data Sets > Data Sets with Detailed Tables > Geography > Tables > Results  
2006 American Community Survey, Detailed Tables

Choose a selection method

list name search address search map geo within geo

Explain Census Geography | Where are Geographic Components (Urban and Rural)?

Select a geographic type

Metropolitan Statistical Area/Micropolitan Statistical Area

Select one or more geographic areas and click 'Add'

- San Francisco-Oakland-Fremont, CA Metro Area
- San Germán-Cabo Rojo, PR Metro Area
- San Jose-Sunnyvale-Santa Clara, CA Metro Area
- San Juan-Caguas-Guaynabo, PR Metro Area
- San Luis Obispo-Paso Robles, CA Metro Area
- Sandusky, OH Metro Area
- Santa Barbara-Santa Maria, CA Metro Area
- Santa Cruz-Watsonville, CA Metro Area

Explain Missing Geographies

Map It

Add

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

Figure 15. **Selecting Variables in American FactFinder**

The screenshot shows the 'Select Tables' page in American FactFinder. The breadcrumb trail is: Main > Data Sets > Data Sets with Detailed Tables > Geography > Tables > Results > 2006 American Community Survey, Detailed Tables. There are three selection methods: 'by subject', 'by keyword', and 'show all tables'. The 'show all tables' method is selected. A list of tables is displayed, with 'B08122. Means of Transportation to Work by Poverty Status in the Past 12 Months' circled in orange. Below the list is an 'Add' button. Underneath, the 'Current table selections:' section shows the selected table name.

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

Figure 16. **Example of Transportation Data Available in American FactFinder**

The screenshot shows the 'Detailed Tables' page for table B08122. The breadcrumb trail is: Main > Data Sets > Data Sets with Detailed Tables > Geography > Tables > Results. The 'Print / Download' link is circled in orange. Below the table title is a data table for 'San Juan-Caguas-Guaynabo, PR Metro Area'.

San Juan-Caguas-Guaynabo, PR Metro Area		
	Estimate	Margin of Error
Total:	787,422	+/-12,668
Below 100 percent of the poverty level	145,787	+/-5,801
100 to 149 percent of the poverty level	134,138	+/-6,836
At or above 150 percent of the poverty level	507,497	+/-13,260
Car, truck, or van - drove alone:	592,808	+/-11,362
Below 100 percent of the poverty level	95,419	+/-4,516
100 to 149 percent of the poverty level	97,498	+/-5,890
At or above 150 percent of the poverty level	399,891	+/-11,306
Car, truck, or van - carpooled:	99,021	+/-4,801
Below 100 percent of the poverty level	17,675	+/-2,035
100 to 149 percent of the poverty level	17,787	+/-1,993
At or above 150 percent of the poverty level	63,559	+/-3,924

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

To calculate poverty rates and their associated margins of error, Virginia first needs to put the data into a spreadsheet. Since this is a small table, she uses her mouse to copy and paste the data directly into a spreadsheet program. Others may prefer to download the data using the *Print/Download option*, as shown in Figure 16.

Next, she needs to derive the poverty rates for those workers who drive to work and those who take public transportation. In Appendix 3 of this handbook, she finds the formulas needed to calculate the margin of error for a derived proportion. She calculates the proportion of workers who drive alone to work for whom poverty status is determined.

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

Where  $\hat{X}_{num}$  is the estimate used in the numerator, the number of workers who drive alone (car, truck, or van—drove alone) and who are below 100 percent of the poverty level (95,419).  $\hat{X}_{den}$  is the estimate used in the denominator, the total number of workers who drive alone (592,808).

$$\hat{p} = \frac{(95,419)}{(592,808)} = 0.1610$$

Next she needs to calculate the margin of error associated with her derived proportion. Again, she uses the appropriate formula from Appendix 3:

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

For the poverty rate of workers that drive alone:

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

$$MOE_p = \frac{\pm \sqrt{(4,516)^2 - ((0.161)^2 * (11,362)^2)}}{592,808}$$

$$MOE_p = \frac{\pm \sqrt{20,394,256 - (0.025921 * 129,095,044)}}{592,808}$$

$$MOE_p = \frac{\pm 4,129}{592,808} = \pm .007$$

Therefore, the proportion of workers that drive alone ( $\hat{p}$ ) is 0.161 with a margin of error ( $MOE_p$ ) of  $\pm 0.0070$ . Applying the same formulas to the PRCS data for workers that take public transportation and multiplying all of the results by 100, Virginia comes up with the data shown in Table 6.

Since the margin of error for workers who take public transportation is relatively large, she may want to use the instructions in Appendix 4 of this handbook to test whether the difference between these two PRCS estimates is statistically significant. Based on her preliminary results, she reports to the other members of the planning commission that the poverty rate for workers that take public transportation to work in San Juan-Caguas-Guaynabo Metro Area, who numbered more than 32,000 just in the past year, may be more than two times higher than the poverty rate for workers that drive alone.

**Table 6. Poverty Rates for Workers That Drive Alone and Workers That Take Public Transportation in San Juan-Caguas-Guaynabo PR Metro Area: 2007**

	Poverty rate	Margin of error (+/-)
Drive Alone	16.09	0.70
Take Public Transportation	36.29	4.56

Source: U.S. Census Bureau, 2007 American Community Survey.

# Avoiding Pitfalls When Working With PRCS Data

---

The PRCS data are complex and cover a broad range of topics and geographic areas. Because this is a relatively new survey, many people do not fully understand how to interpret and use the PRCS data. Some key points are summarized below to guide you as you explore the PRCS data.

- Use caution in comparing PRCS data with data from the decennial census or other sources. Every survey uses different methods, which could affect the comparability of the numbers.
- The PRCS was designed to provide estimates of the characteristics of the population, not to provide counts of the population in different geographic areas or population groups.
- Be careful in drawing conclusions about small differences between two estimates because they may not be statistically different.
- Data users need to be careful not to interpret annual fluctuations in the data as long-term trends.
- Use caution in comparing data from 2006 and later surveys with data from the 2005 PRCS. Unlike the 2005 survey, the 2006 PRCS survey includes samples of the population living in group quarters (e.g., college dorms and nursing homes), so the data may not be comparable, especially for young adults and the elderly who are more likely than other age groups to be living in group quarters facilities.
- Data users should not interpret or refer to 3-year or 5-year period estimates as estimates of the middle year or last year in the series. For example, a 2005–2007 estimate is not a “2006 average.”
- Data users should not rely on overlapping confidence intervals as a test for statistical significance because this method will not always provide an accurate result.

## Other Resources for Working With ACS and PRCS data

---

Data users can find a wealth of additional information about the ACS/PRCS on the Census Bureau’s Web site at <[www.census.gov/acs](http://www.census.gov/acs)>. However, sorting through all of the information available on the ACS/PRCS can take a long time without a good map of the available

resources. In Appendix 8, we provide links to online resources that data users are likely to find the most useful. These resources cover many of the topics discussed in this handbook, but in greater detail.

## ACS/PRCS Data Resources

- FactFinder Help (online help, census data information, glossary, and tutorial)  
<<http://factfinder.census.gov/home/en/epss/main.html>>
- 2007 Guide to the Data Products (Web page)  
<[http://www.census.gov/acs/www/Products/users\\_guide/](http://www.census.gov/acs/www/Products/users_guide/)>
- Using Multiyear Estimates (Web page)  
<<http://www.census.gov/acs/www/UseData/myeoverview.html>>
- American Community Survey Design and Methodology (technical paper)  
<<http://www.census.gov/acs/www/Downloads/tp67.pdf>>
- Accuracy of the ACS Data (online document)  
<<http://www.census.gov/acs/www/Downloads/ACS/accuracy2007.pdf>>
- ACS Sample Size (Web page)  
<<http://www.census.gov/acs/www/SBasics/SSizes/SSizes07.htm>>
- ACS Quality Measures (Web page)  
<<http://www.census.gov/acs/www/UseData/sse/>>
- 2006 Data Users Handbook: PRCS (online document)  
<[http://www.census.gov/acs/www/Downloads/2006%20PRCS%20Data%20Users%20Handbook%20\(English\).pdf](http://www.census.gov/acs/www/Downloads/2006%20PRCS%20Data%20Users%20Handbook%20(English).pdf)>
- How to Use the ACS Data (Web page)  
<<http://www.census.gov/acs/www/UseData/>>
- Guidance on Comparing 2007 ACS Data (online tables)  
<<http://www.census.gov/acs/www/UseData/compACS.htm>>
- Using Different Sources of Data for Income and Poverty (online fact sheet)  
<<http://www.census.gov/hhes/www/income/factsheet.html>>

## Conclusion

Data from the PRCS can be extremely valuable for a variety of different purposes that include: to monitor the well-being of Puerto Rico's children and families, to investigate the characteristics of the workforce, to determine the economic well-being of working-poor families, or to track social, economic, and demographic changes in the general population. Many people are

being cautious in their approach to the PRCS, and rightly so. This is a relatively new survey with a new approach to measuring change in our communities. The PRCS has great potential, particularly as a source of annual data for local areas. By providing data for communities each year, the PRCS will provide critical information for communities when they need it most.

# 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 the 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 Multi-year 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.

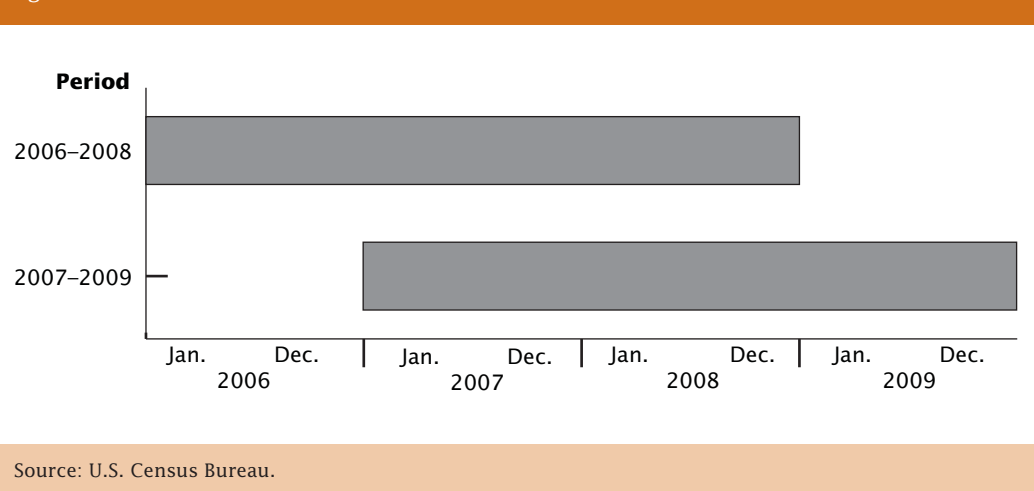
You certainly can compare characteristics for areas with populations of 30,000 to areas with populations of 100,000 but you should use the data set that they have in common. In this example you could use the 3-year or the 5-year estimates because they are available for areas of 30,000 and areas of 100,000.

## 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

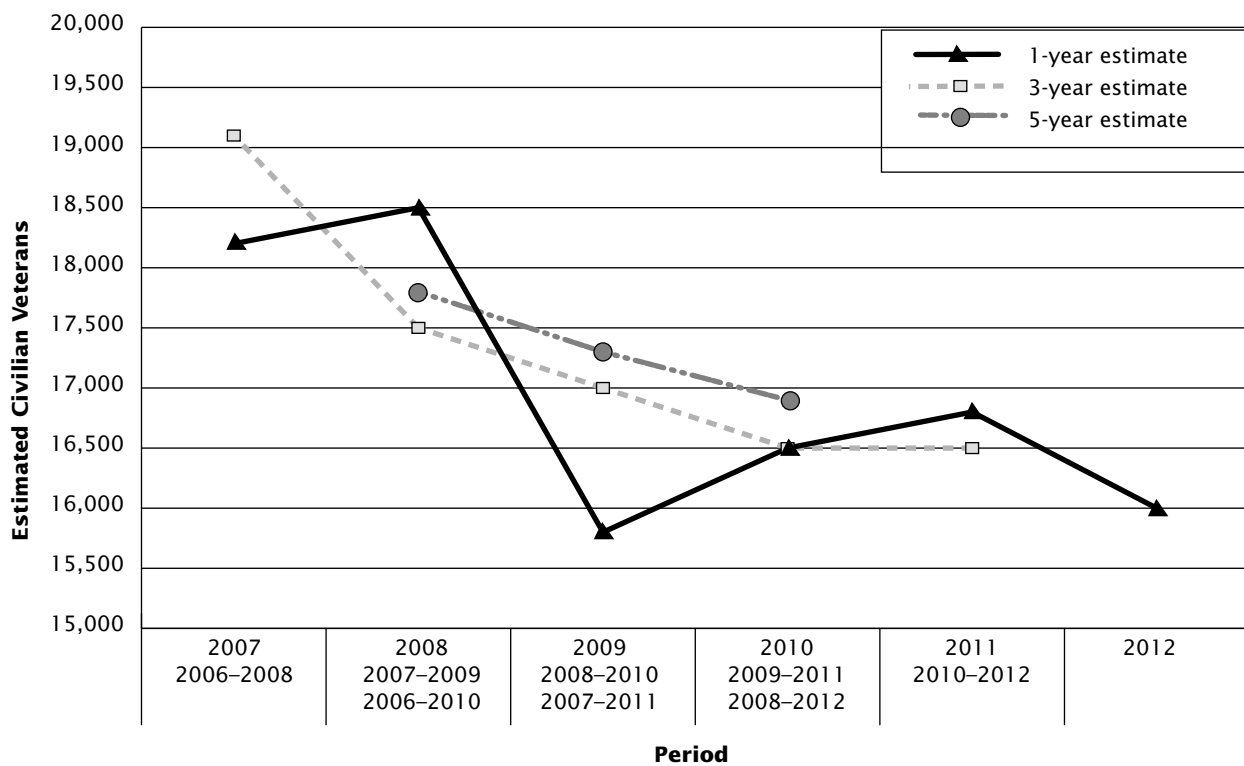


Variability in single-year estimates for smaller areas (near the 65,000-publication threshold) and small sub-groups within even large areas may limit the ability to examine trends. For example, single-year estimates for a characteristic with a high CV vary from year to year because of sampling variation obscuring an underlying trend. In this case, multiyear estimates may be useful for assessing an underlying, long-term trend. Here again, however, it must be recognized that because the multiyear estimates have an inherent smoothing, they will tend to mask rapidly developing changes. Plotting the multiyear estimates as representing the middle year is a useful tool to illustrate the smoothing effect

of the multiyear weighting methodology. It also can be used to assess the “lagging effect” in the multiyear estimates. As a general rule, users should not consider a multiyear estimate as a proxy for the middle year of the period. However, this could be the case under some specific conditions, as is the case when an area is experiencing growth in a linear trend.

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

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



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

## Summary of Guidelines

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

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

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

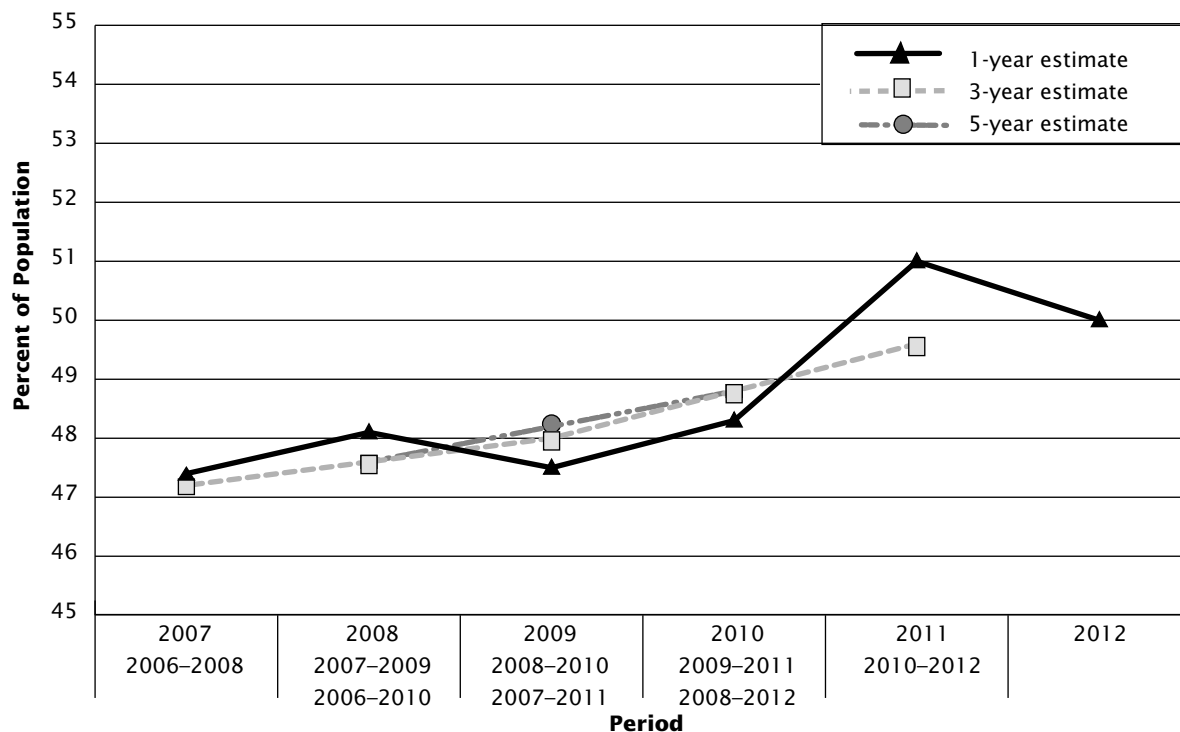
ated with sufficiently small CVs, can be very valuable in identifying and studying such phenomena. Graphing trends for such areas using single-year, 3-year, and 5-year estimates can take advantage of the strengths of each set of estimates while using other estimates to compensate for the limitations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

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

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

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

Figure 3. **Proportion of Population With Bachelor's Degree or Higher, City X Single-Year, Multiyear Estimates**



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

## Appendix 2.

# Differences Between ACS and Decennial Census Sample Data

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

### Residence Rules, Reference Periods, and Definitions

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

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

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

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

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

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

### More on Residence Rules

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

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



count of the population as of Census Day, April 1, while the ACS residence rules were designed to collect representative information to produce annual average estimates of the characteristics of all kinds of areas. When interviewing the population living in housing units, the decennial census uses a “usual residence” rule to enumerate people at the place where they live or stay most of the time as of April 1. The ACS uses a “current residence” rule to interview people who are currently living or staying in the sample housing unit as long as their stay at that address will exceed 2 months. The residence rules governing the census enumerations of people in group quarters depend on the type of group quarter and where permitted, whether people claim a “usual residence” elsewhere. The ACS applies a straight de facto residence rule to every type of group quarter. Everyone living or staying in a group quarter on the day it is visited by an ACS interviewer is eligible to be sampled and interviewed for the survey. Further information on residence rules can be found at <http://www.census.gov/acs/www/AdvMeth/CollProc/CollProc1.htm>.

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

### More on Reference Periods

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

represent the average characteristics over a full year (or sets of years), a different time, and reference period than the census.

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

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

### Income Data

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

### School Enrollment

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

### Utility Costs

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

### Definitions

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

---

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

### **Implementation**

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

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

# 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.<sup>1</sup> 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  $\pm 12,357$ . 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.<sup>2</sup>

Derivation of SEs can thus be expressed as

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

<sup>1</sup> 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.

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

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  $\pm 12,357$ . 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 variabilities. 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  $\pm 1,645$ , 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

$$CI_{CL} = (L_{CL}, U_{CL}).^3$$

<sup>3</sup> 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 ( $\pm 14,723$ ).

$$L_{95} = 771,782 - 14,723 = 757,059$$

$$U_{95} = 771,782 + 14,723 = 786,505$$

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 ( $[1,000 / 20,000] \times 100$ ). 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)

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

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	$\pm 3,303$
Females living alone in Arlington County (Component 2)	19,464	$\pm 2,011$
Females living alone in Alexandria city (Component 3)	17,190	$\pm 1,854$

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  $\pm 4,289$ .

### 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.

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

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.

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

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  $50,624 \times 0.134 = 6,784$ . The MOE is calculated by:

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

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

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

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:

$$\begin{aligned} \text{Percent Change} &= 100\% * \frac{\hat{X}_2 - \hat{X}_1}{\hat{X}_1} = \\ 100\% * \left( \frac{802,210 - 762,475}{762,475} \right) &= 5.21\% \end{aligned}$$

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 ( $MOE_1$ ):

$$\begin{aligned} MOE_2^2 + (\hat{R}^2 * MOE_1^2) &= \\ (22,866)^2 + [(1.0521)^2 * (22,666)^2] &= \\ 1,091,528,529 \end{aligned}$$

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).<sup>4</sup>
- 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.

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

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

Algebraically, the significance test can be expressed as follows:

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

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

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

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

$Z_{CL}$  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:

$$\begin{aligned} \left| \frac{\hat{X}_1 - \hat{X}_2}{\sqrt{SE_1^2 + SE_2^2}} \right| &= \left| \frac{22.0 - 21.5}{\sqrt{(0.12)^2 + (0.12)^2}} \right| = \\ \left| \frac{0.5}{\sqrt{0.015 + 0.015}} \right| &= \left| \frac{0.5}{\sqrt{0.03}} \right| = \left| \frac{0.5}{0.173} \right| = 2.90 \end{aligned}$$

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

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

$$\left| \frac{\hat{X}_1 - \hat{X}_2}{\sqrt{SE_1^2 + SE_2^2}} \right| = \left| \frac{22.5 - 22.1}{\sqrt{(0.20)^2 + (0.25)^2}} \right| = \left| \frac{0.4}{\sqrt{0.04 + 0.0625}} \right| = \left| \frac{0.4}{\sqrt{0.1025}} \right| = \left| \frac{0.4}{0.320} \right| = 1.25$$

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

### Comparisons Within the Same Time Period

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

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

### Comparisons Across Time Periods

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

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

### Comparisons Based on Overlapping Periods

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

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

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

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

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

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

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

### Comparisons With Census 2000 Data

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

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

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

<sup>5</sup> 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](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.

# 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](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:<sup>6</sup>

- 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.

<sup>6</sup> 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.

# 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](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.

