Speaker Notes:

Data Products from the American Community Survey
Updated: 2/20/09

Slide 1

The American Community Survey, also called the ACS, is a new approach for collecting accurate, timely information needed for critical government functions. This new approach provides accurate, up-to-date profiles of America's communities every year. The American Community Survey provides annual estimates of demographic, housing, social, and economic characteristics for numerous geographies every year. The American Community Survey provides one-year estimates for all states as well as for cities, counties, and metropolitan areas with a total population of 65,000 or more. Beginning in December 2008, the American Community Survey will provide three-year estimates for geographies with a total population of 20,000 or more. And by the end of 2010, the American Community Survey plans to provide 5-year estimates for all geographies, even those with very small populations.

Slide 2

There are several key differences between the data produced from the American Community Survey and the data produced from Census 2000 and earlier censuses. These differences include the concept of period estimates and the publication of sampling error measures in many of the ACS data products. We'll briefly look at these two issues before getting into the data products.

Slide 3

The main function of the decennial census is to provide counts of every person residing in the United States for the purposes of Congressional apportionment and legislative redistricting. The primary purpose of the American Community Survey is to provide a portrait of the social, housing, economic, and demographic characteristics of the U.S. population. As a result, the American Community Survey does not provide official counts of the population or housing. In between censuses the Census Bureau's Population Estimates Program is the official source for annual housing unit totals and for annual population totals by age, sex, race, and Hispanic origin.

The basic rule is that if you need a population count, use decennial census data or data from the Population Estimates Program. If you need to describe the characteristics of the population, use data from the American Community Survey.

Slide 4

The American Community Survey 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 American Community Survey one-year estimates, the period is the calendar year. For example, the 2007 American Community Survey data describe the population and housing characteristics of an area from January 1, 2007 through December 31, 2007, not for any specific day within the year.

A period estimate is different from a point-in-time estimate. A point-in-time estimate is designed to measure characteristics as of a certain date or narrow time period. For example, the purpose of the decennial census is to count the population living in the United States on a specific date, which is traditionally April 1. Although decennial census data are actually collected over several months, they are designed to provide a snapshot of the U.S. population as of April 1.

Slide 5

Sampling error is the uncertainty associated with an estimate that is based on data gathered from a **sample** of the population rather than the **full** population.

So, why do sample estimates have uncertainty associated with them? There are two reasons:
Estimates of characteristics from the sample data can differ from those that would be obtained if the entire population were surveyed.
Estimates from one subset or sample of the population can differ from those based on a different sample from that same population.

A key measure of sampling error is the margin of error. It is defined as a measure of the precision of an estimate at a given level of confidence. The most commonly used confidence levels are 90%, 95% and 99%. So what does the confidence level of a margin of error mean? The confidence level of a margin of error indicates the likelihood that the difference between the population value and the sample estimate is less than or equal to the margin of error.

American Community Survey estimates are published with their margins of error at the 90 percent confidence level. However, it is possible to construct margins of error with higher levels of confidence, such as 95 percent or 99 percent. This is done by adjusting the published margin of error. Instructions for these adjustments can be found in the technical appendices in the Compass Handbooks available on the Census Bureau's web site.

It is important to note that the long form data from Census 2000 are also sample data. Therefore estimates produced from the long form also had sampling error associated with them, but the Census Bureau did not publish these data within the data products.

Slide 6

The Census Bureau releases a suite of American Community Survey data products every year. The data products are similar to those produced for Census 2000.

Slide 7

The American Community Survey data products are similar to those produced from the decennial census long form. Like the decennial sample data products, the American Community Survey products show the characteristics of the country's population and housing.
These products include four broad types of products – profiles, tables, thematic maps, and Public Use Microdata Sample, or PUMS, files. There are multiple types of profiles and tables – data profiles, narrative profiles, comparison profiles, selected population profiles, detailed tables, subject tables, ranking tables, and geographic comparison tables.

American Community Survey data are available on the Census Bureau's American FactFinder.

The upcoming slides will go over each of these products in detail. Many of these slides will show static images of the data products from the 2007 American Community Survey taken in late-September 2008. Later in this presentation we will go through an example of how to access the American Community Survey data products.

Slide 8

Data Profiles are a good place for novice data users to start. They provide separate fact sheets on the social, economic, housing, and demographic characteristics for different geographic areas. The four profiles include a total of about 450 different characteristics. Data Profiles can be displayed in either a tabular or narrative format.

This slide shows an image of a tabular data profile. The tabular version is a five-column table that displays a given set of selected summary characteristics and other derived measures for each geographic area in the first column on the left. The second and third columns include the estimate and its associated margin of error, respectively. The fourth and fifth columns display the estimates in percentage form and the margins of error associated with those percentages, respectively.

As with all American Community Survey data products, the margins of error are presented at the 90-percent confidence level.

Slide 9

Narrative profiles provide clear, concise textual descriptions of the data included in the tabular data profiles. These narratives are easy-to-read, computer-produced profiles that summarize information on a wide array of subjects in words, rather than numbers, for novice data users.

The image on the slide shows part of a typical narrative profile. Two measures are summarized in this image – households and families and nativity and language. The households and families measure is summarized in both narrative format and a simple bar chart that displays the types of households in the United States in 2007. The nativity and language measure is summarized in narrative format only. Let's read this narrative so we can see how the narrative data profiles summarize information:

Thirteen percent of the people living in United States in 2007 were foreign born. Eighty-seven percent was native. Of those born in United States, 59 percent was born in their state of residence.

Among people at least five years old living in United States in 2007, 20 percent spoke a language other than English at home. Of those speaking a language other than English at home, 62 percent spoke Spanish and 38 percent spoke some other language; 44 percent reported that they did not speak English "very well."

Slide 10

Comparison Profiles show data side-by-side from the 2006 American Community Survey and the 2007 American Community Survey, indicating where there is a statistically significant difference between these two sets of estimates. An * in the statistical significance column of the comparison profile indicates that the 2006 ACS estimate is significantly different from the 2007 ACS estimate. A “c” indicates both estimates are controlled and thus a test of statistical significance is not appropriate. The comparison profiles are only available for one-year estimates. There are four types of comparison profiles – social, economic, housing, and demographic and they display the same characteristics as the data profiles.

The slide displays an image of a portion of the social comparison profile. The first of this four-column table contains the measures for which data are presented. The second column displays data from the 2007 American Community Survey and the third column displays data from the 2006 American Community Survey. The fourth column displays the result of the statistical significance test. For example, the 15th row shows information on households with one or more people under 18 years. In 2007, this group made up an estimated 34.4 percent of all households in the United States. In 2006, households with one or more people under 18 years made up an estimated 34.6 percent of all households in the United States. There is an * in the statistical significance column, indicating that these estimates are statistically different from one another. These estimates show that there was a change between the 2007 and 2006 American Community Survey estimates that is statistically significant at the 90 percent confidence level.

Slide 11

The most detailed race and ethnic data are available through the Selected Population Profiles, which provide summary tables separately for over 100 detailed race, ethnic, ancestry, and tribal groups. Starting with the 2007 American Community Survey, the selected population profiles are also available for approximately 100 countries and region of birth groups. The slide displays three images from the American FactFinder population group selection page. The three images correspond to the three selection tabs that are available: Race or Ethnic Groups, Ancestry Groups, and Country of Birth. The first ten options in each category are shown in the images.

There are two thresholds used for Selected Population Profiles. First, we look at the total population size of the geographic area. Then we look at the size of the population group of interest. In the 2007 American Community Survey, Selected Population Profiles are only released for geographic areas

with a total population of at least 500,000. This threshold allows the release of Selected Population Profiles for all congressional districts. Selected Population Profiles that are based on one-year estimates are only released for selected population groups with a minimum estimated population size of 65,000. The three-year Selected Population Profiles will support selected populations with a minimum estimated population size of 20,000. Four new estimates were added to the Selected Population Profile identifying the number of races reported.

For example, suppose we're interested in the Chinese population in San Francisco County, California. In 2007, San Francisco County had a total population of 764,796 people and there were an estimated 154,096 people who reported their race as Chinese alone. Because this situation meets both the geographic area population threshold and the population group of interest population threshold, a Selected Population Profile was produced.

Slide 12

Selected population profiles provide the user with a ready-made report on a population group and includes the 90 percent margin of error. The idea is to quickly produce a report on a population group of interest, for example: Native Hawaiians and Other Pacific Islanders, people of Bulgarian descent, or the foreign-born population born in Brazil.

This slide displays a portion of the selected population profile for the foreign-born population in the United States who reported that they were born in Japan. The first column of the five-column table displays selected characteristics for which data are displayed. The second and third columns show the estimates and margins of error for the total population of the United States, respectively. The fourth and fifth columns show the estimates and margins of error for the selected population group, respectively, in this instance all foreign-born people in the United States who reported that they were born in Japan.

Slide 13

The detailed tables provide the most detailed data on all topics and geographic areas and are the foundation upon which other data products are built.

Detailed Tables are basic distributions of characteristics that show estimates and their associated margins of error. There are more than 1,200 Detailed Tables and many of the tables are repeated for 11 race and Hispanic origin groups. Detailed tables include distributions for more than 500 characteristics, over 300 race and Hispanic Origin iterations, and 81 imputation tables.

A few examples of the types of topics covered by our detailed tables…
Sex by Age by Race and Hispanic Origin
Means of Transportation to Work by Travel Time to Work
Median Number of Rooms in Housing Units
School Enrollment by Level of School
Poverty Status in the past 12 Months by Sex and Age

Slide 14

This slide displays an example of a detailed table as displayed on the American FactFinder. The table is B16007. Age by Language Spoken at Home for the Population 5 Years and Over from the 2007 American Community Survey 1-Year Estimates.

The table title at the top describes the variables in the table, and any combination of them for which estimates are presented and the universe for which the estimates apply. The data set indicates the year the data were collected and the type of estimate.

The table has an additional column containing the margin of error for the 90 percent confidence level of the estimate. Confidence bounds can be created by adding the margin of error to the estimate (for an upper bound) and subtracting the margin of error from the estimate (for a lower bound). All published margins of error for the American Community Survey are based on a 90 percent confidence level.

Due to the level of specificity of the detailed tables, many tables have a "collapsed" version. If a full detailed table has a collapsed version, it will be indicated at the top of the table.

Slide 15

Collapsed tables show less detailed information – at least two of the lines from the original detailed table have been collapsed into one line.
This is especially useful for smaller geographic areas that might not be able to populate enough of the table cells for the table to pass data quality filters.

For example, this table is C16007, which is the collapsed version of table B16007 that we looked at on the previous slide. The detail in the red box shows the original categories. Speak other Indo-European languages and Speak Asian and Pacific Island languages have been combined with Speak other languages to form a single category.

Collapsed tables are noted by the letter "C" as the first character in the table number. When selecting a table from the American FactFinder, collapsed tables immediately follow their source detailed table and the entry is in blue as opposed to black.

Slide 16

Using the "options" tab in American FactFinder, you can display information by different geographic components for the detailed and collapsed tables. Geographic components are available for the nation, region, divisions, and states. Fifteen geographic components are available from the American Community Survey, including urban, rural, in metropolitan or micropolitan statistical area, not in metropolitan or micropolitan statistical area, and several others.

In this slide, we started out by selecting California as our initial geography. Using the options tab found beneath the American FactFinder banner, we selected urban and rural as the geographic

components we wanted displayed. This results in a table showing three different geographies – California, Urban California, and Rural California. The estimates and margins of error are displayed for each of the three geographies.

Slide 17

Quality measures illustrate the steps the Census Bureau takes to ensure that American Community Survey data are accurate and reliable. Beginning with the 2007 American Community Survey, the quality measures are available through American FactFinder in the B98 series of Detailed Tables. There are ten tables that describe the quality of the American Community Survey sample and the data it collects, including sample sizes, coverage rates, and response rates.

Slide 18

For information about a particular topic, such as employment, education, and origins and language, users should start with the Subject Tables. Subject Tables provide pretabulated numbers and percentages for a wide variety of topics, often available separately by age, sex, race or ethnicity.

The image on the slide displays a partial list of the available Subject Tables in the American FactFinder. Subject Tables are largely derived from the detailed tables and are similar to the Census 2000 Quick Tables.

Slide 19

Subject tables show more detail than is available in the data profiles. Generally, they present distributions for a few key population groups, with universes displayed as numeric estimates with their margins of error. Subject tables display measures such as medians and aggregates where appropriate, and include tables of imputation rates for relevant measures. The American FactFinder contains over 60 summarized topic-specific subject tables.

In the example displayed on the slide, we are looking at table S1401 School Enrollment for California. This is a seven-column table displaying information on school enrollment for the total population, the population enrolled in public school and the population enrolled in private school.

Slide 20

Ranking Tables provide state-level rankings of key ACS variables. There are more than 80 ranking tables available on various topics. The image on the slide shows ranking table R1601. Percentage of People 5 Years and Over Who Speak a Language Other Than English at Home. This is a four-column table, showing the rank in the first column, the state in the second, and the estimate and margin of error in the third and fourth columns, respectively. Each table displays a single characteristic and the states are “ranked” from highest to lowest. This enables users to compare the different states while seeing which differences among the states are statistically significant at a 90-

percent confidence level. This can be seen by clicking on “with statistical significance,” found in the menu on the left side of the screen.

Slide 21

When viewing a table with statistical significance, the user can select what state they are interested in. This is done either by clicking directly on the state of interest or by using the drop down menu above the table. Ranking tables shown with statistical significance are five column tables. A thin column is inserted between the “rank” and “state” columns, which is where the statistical significance indicator is displayed.

In this slide, the user selected New Jersey, as evidenced by the two dots in the column to the right of “rank.” The presence of a single dot indicates states whose estimates are not statistically different from the estimate of the selected geography. So for this example, it would be incorrect to draw conclusions based on a comparison of the estimates for New Jersey and Nevada.

Slide 22

Users interested in geographic comparisons for areas other than states may be interested in the Geographic Comparison Tables. These tables allow comparisons of ACS data across a variety of geographic areas, including metropolitan areas, cities, counties, and Congressional districts.

This slide shows the percent of people 5 years and over who speak English less than “very well” broken out by the urban and rural areas of California as well as inside and outside metropolitan and micropolitan statistical areas. This is a three column table that displays the geographic area in the first column, the estimate in the second column, and the margin of error in the third column.

Slide 23

Thematic Maps provide graphic displays of the data available through the geographic comparison tables and ranking 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 state, county, metropolitan area, or Congressional district of interest.

In this slide, the red pin indicates that Maine is the selected geography and the estimates for the states with hash-marks do not have significant statistical differences from the estimate for Maine. As the legend on the left explains, the lighter shades indicate lower values and darker shades indicate higher values.

Slide 24

Those with expertise in using SAS, SPSS, STATA, or Data Ferrett may also be interested in the Public Use Microdata Sample, or PUMS, files. These files contain a subsample of individual

records of people and housing units that responded to the survey. All identifying information, such as name and address, have been stripped from each individual record to ensure confidentiality. The records contain information, on the selected subsample of housing units and group quarter persons, captured in their completed ACS questionnaires. The questionnaire includes questions on age, sex, tenure, income, education, language spoken at home, journey to work, occupation, condominium status, shelter costs, vehicles available, and other subjects.

For many data users, the summary tables and tabular and narrative profile reports will suffice. Microdata are for those users who want to create do-it-yourself tabulations, to be able to further draw on the richness of detail recorded in the ACS survey. The PUMS files permit analysis of specific population groups and custom variables that are not available through the American FactFinder.

Microdata users frequently want to look at relationships among variables not shown in the standard products offered by the Census Bureau. For example, what are the characteristics of unemployed homeowners? What characteristics do families with four or more children have in common? What kinds of Hispanic families in a state own their own homes?

The advantage of PUMS is that data users can tabulate data according to the characteristics they need to know about. The disadvantage of PUMS is that data are only available for states and large geographies called Public Use Microdata Areas or PUMAs. When using the PUMS files, users are sacrificing geographic detail for subject detail.

Slide 25

There are several ways in which you can access ACS data products.

Slide 26

All ACS data products are available on the American FactFinder and via the FTP site for the ACS. You can also access the PUMS files through the DataFerrett. Let's look at each method in a bit more detail.

To access ACS data products through the American FactFinder, begin by visiting the main census website: www.census.gov

On the left menu of the home page, the third link from the top is a link to the American FactFinder, which is the main vehicle for accessing the data products.

Slide 27

There are various links on the introductory page for the American FactFinder. In the middle of the page is a column called "getting detailed data." There are several data sources listed under this

column, including the American Community Survey. We can get to the ACS data by clicking on the link that says “get data.”

Slide 28

Once you have clicked on “get data” you will come to a page that displays all of the available years of data for the American Community Survey. The most recent release of data is automatically selected as the default dataset. To change to an earlier year, you must select the radio button next to the year you are interested in.

Once you have selected the year that you are interested in, on the right you will notice a list of all of the available data products for that year. From this list you can select the type of data product you are interested in retrieving. From this page you can also access the PUMS files. This access point is located in the box in the upper right of the screen called “Other Resources.”

Say, for instance, we want to generate the Geographic Comparison Table that we looked at earlier. But instead of looking at California broken out by urban and rural as well as metropolitan and micropolitan statistical areas, we want to look at the percent of people who speak English less than “very well” in California by county. The first step in generating this table is to click on “Geographic Comparison Tables.”

Slide 29

To create a Geographic Comparison Table we first need to select the geography of interest.

For our example, we will select “State” under type, which brings us to the next geographic selection where we select “California.” The next box contains options for creating a Geographic Comparison Table. You’ll see that we can compare geographies other than counties, but for this example we will select “State – County.” Then we click “NEXT.”

Slide 30

This brings us to the “Table” screen. Here we select our subject of interest. Because we are interested in comparing the percent of people who have some difficulty speaking English, we highlight table GCT1603 “Percent of People Who Speak English Less Than ‘Very Well.’” To retrieve the data table, we click on “Show Result.”

Slide 31

This slide shows the resulting Geographic Comparison Table. This table displays the percent of people who speak English less than “very well” for all California counties with populations of 65,000 or more. Due to the number of counties in California, not all of the counties could be displayed on this slide.

For other useful tips on using the data products you can visit the 2007 Data Product Details web page, which can be found on the American Community Survey website.

Slide 32

Additional user assistance can be found at the main page of American FactFinder. Clicking on the "Help" tab at the top of the page will open a window that will point you to a set of tutorials on how to use the American FactFinder.

Slide 33

You can also access ACS data through the American Community Survey FTP Site. The FTP site can be accessed via the address listed on the slide or through the ACS website, which you can access from the main Census.gov page. The FTP site contains ACS data as they are released. On the FTP site users can find current and historical ACS data. The FTP site is particularly useful for users who want to download all of the tables for a specific geography.

For example, if you are a local planner in Bucks County, Pennsylvania, it may be beneficial to use the FTP site to download all of the detailed tables for Bucks County.

Slide 34

As mentioned before, the PUMS files can also be accessed through a data access tool called the DataFerrett. DataFerrett is available on the Census website and can be downloaded for free. DataFerrett is a data analysis tool with the ability to access many federal, state, local, and private datasets via an Internet connection. For people who do not have access to data analysis tools such as SAS, SPSS, or STATA, the DataFerrett is an easy way to access the PUMS files and create the tables you are interested in.

Slide 35

You can download and install DataFerrett to your computer from the website listed on the slide. DataFerrett allows you to recode variables and create tabulations from survey microdata. Tutorials on using DataFerrett are available within the application. You can access them from the lower right corner of the main DataFerrett web page. It is highly recommended that you review these before using DataFerrett, as there are many types of analyses that DataFerrett is capable of computing.

Slide 36

This presentation gave you an overview of the data products available from the American Community Survey.

The American Community Survey staff has developed the ACS Alert, which is an e-mail newsletter giving data users the latest news about the survey. You can subscribe to the newsletter by contacting the American Community Survey staff or read past editions of the "ACS Alert" on the Internet at: http://www.census.gov/acs/www/Special/Alerts.htm

Please feel free to contact the Census Bureau if you have questions or need further information. If you have questions that are not answered by the Web site, please call 1-800-923-8282 or email acso.users.support@census.gov.