ACS Advanced Applications and Issues

Working with ACS data requires an appreciation for the advantages and disadvantages of single year and multiyear estimates, the ability to determine and interpret sampling error measures associated with individual estimates, and an understanding of the issues affecting use. These appendices are intended to supplement the ACS Handbooks by providing the practitioner with information concerning single year and multiyear estimates and guidelines for their use, along with information concerning sampling errors and methods for calculating alternative sampling error measures and for using sample error measures when comparing estimates.

The following Appendices are available:

<table>
<thead>
<tr>
<th>#</th>
<th>Description</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Understanding ACS Single Year Estimates</td>
<td>2-3</td>
</tr>
<tr>
<td>2</td>
<td>Understanding ACS Multiyear Estimates</td>
<td>4-6</td>
</tr>
<tr>
<td>3</td>
<td>Differences Between ACS and Decennial Census Sample Data</td>
<td>7-8</td>
</tr>
<tr>
<td>4</td>
<td>Sampling Errors</td>
<td>9-11</td>
</tr>
<tr>
<td>5</td>
<td>Deriving Sampling Error Measures</td>
<td>12-14</td>
</tr>
<tr>
<td>6</td>
<td>Guidance on Use of ACS Single and Multiyear Estimates</td>
<td>15-20</td>
</tr>
<tr>
<td>7</td>
<td>Calculating Margins of Error for Derived Estimates</td>
<td>21-25</td>
</tr>
<tr>
<td>8</td>
<td>Making Comparisons</td>
<td>26-32</td>
</tr>
<tr>
<td>9</td>
<td>Using Dollar-denominated Data</td>
<td>33-35</td>
</tr>
<tr>
<td>10</td>
<td>Quality Measures</td>
<td>36-39</td>
</tr>
<tr>
<td>11</td>
<td>Implication of Population Controls on ACS Estimates</td>
<td>40-40</td>
</tr>
<tr>
<td>12</td>
<td>Other ACS Resources</td>
<td>41-42</td>
</tr>
</tbody>
</table>
Appendix 1. Understanding ACS Single Year Estimates

The ACS produces period estimates of socio-economic and housing characteristics, and is designed to provide estimates that describe those characteristics over the full time period. In the case of ACS single year estimates, the period is the calendar year (e.g., 2007 ACS covers January through December 2007).

While one may think of these estimates as representing average characteristics over the 12 months of the calendar year, it must be remembered that the estimates are not directly calculated as an average of twelve monthly values. Rather, the ACS collects survey information continuously nearly every day of the year, and then aggregates the results over the entire year. The data collection is intended to be evenly spread across the calendar year so as to not over-represent any particular month within the year.

ACS single year estimates provide information about the characteristics of the population and housing for an area over the calendar year. The ACS single year estimates thus contrast with “point-in-time” estimates, such as those from the decennial census long form and 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 U.S. based upon data collected around April 1, 2000, and thus reflects a narrower timeframe than ACS, and CPS collects data for a specific, narrow timeframe (the week containing the 12th of the month) within each month.

For most areas, which have consistent population characteristics throughout a calendar year, the period estimates may not look much different than if the estimates had been obtained from a point-in-time survey design. However, some areas may experience changes in the population such that the characteristics of the population would vary depending upon when in the calendar year measurement occurred. These changes may be due to dynamic changes in the area or its population, or to normal seasonal fluctuations in the population and its associated characteristics across the year. For these areas, the ACS period estimates may look different than if estimates had been obtained from a point-in-time survey design. The impact will be more noticeable for smaller areas (where seasonal changes and dynamic changes such as a factory closing can have a large impact on population characteristics) and for areas with a large impact (for example, the impact in the New Orleans area due to Hurricane Katrina).

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 whose population is dominated by students at a local college which has many more students in the fall and spring semesters than in the summer. While the percent of population enrolled in college or graduate school across the entire year is 50
percent, a point-in-time estimate for any month would yield estimates of either 65 or 20 percent.

<table>
<thead>
<tr>
<th>Month</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan</td>
<td>65</td>
</tr>
<tr>
<td>Feb</td>
<td>65</td>
</tr>
<tr>
<td>Mar</td>
<td>65</td>
</tr>
<tr>
<td>Apr</td>
<td>65</td>
</tr>
<tr>
<td>May</td>
<td>20</td>
</tr>
<tr>
<td>Jun</td>
<td>20</td>
</tr>
<tr>
<td>Jul</td>
<td>20</td>
</tr>
<tr>
<td>Aug</td>
<td>65</td>
</tr>
<tr>
<td>Sep</td>
<td>65</td>
</tr>
<tr>
<td>Oct</td>
<td>65</td>
</tr>
<tr>
<td>Nov</td>
<td>65</td>
</tr>
<tr>
<td>Dec</td>
<td>50</td>
</tr>
</tbody>
</table>

The important thing to keep in mind is that the 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.
Appendix 2. Understanding ACS Multiyear Estimates

The ACS produces period estimates, and is designed to provide estimates that describe characteristics over the full time period. In the case of ACS multiyear estimates, the period is either three or five calendar years (e.g., 2005-2007 ACS covers January 2005 through December 2007, or 2006-2010 ACS covers January 2006 through December 2010).

While one may think of these estimates as representing average characteristics over the 36 or 60 months of the reference period, it must be remembered that the estimates are not calculated as an average of monthly or yearly values. Rather, the ACS collects survey information continuously nearly every day of every year, and then aggregates the results over the entire time period. The data collection is intended to be evenly spread across the entire period so as to not over-represent any particular month or year within the period.

ACS multiyear estimates provide information about the characteristics of the population and housing for an area over the entire multiyear period. The ACS multiyear estimates thus contrast, even more so than ACS single year estimates, with “point-in-time” estimates, such as those from the past decennial census samples and 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 U.S. based upon data collected around April 1, 2000, and thus reflects a narrower timeframe than ACS, and CPS collects data for a specific, narrow timeframe (the week containing the 12th of the month) within each month.

The ACS multiyear estimates also contrast with ACS single year estimates which, while also designed to describe characteristics for a time period, encompass a smaller time period. As discussed later in this Appendix, the differences in time periods between single year and multiyear ACS estimates affect decisions as to which period should be used for a particular analysis.

For most areas, which have consistent population characteristics throughout a calendar year and across time, the period estimates may not look much different than estimates obtained from a point-in-time survey design. However, some areas may experience changes in the population such that the characteristics of the population would vary depending upon when in the period measurement occurred. This is especially true when looking at ACS five-year estimates. These changes may be due to dynamic changes in the area or its population, or to normal seasonal fluctuations in the population and its associated characteristics across the year. For these areas, the ACS period estimates may look different than if estimates had been obtained from a point-in-time survey design. The impact will be more noticeable for smaller areas (where seasonal changes and dynamic changes such as a factory closing can have a large impact.
on population characteristics) and for areas with a large impact (for example, an area with large immigration).

Multiyear estimates are released every year, beginning in 2008 for three-year estimates and 2010 for five-year estimates. The multiyear estimates released in consecutive years thus contain some overlapping years. As shown in Table 1, consecutive three-year estimates contain two years of overlapping coverage, and consecutive five-year estimates contain four years of overlapping coverage. This overlap will affect any comparisons made between consecutive years’ multiyear estimates, and therefore such comparisons should not be made.

### Table 1

<table>
<thead>
<tr>
<th>Data Product</th>
<th>Year of Data Release</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2008</td>
</tr>
<tr>
<td></td>
<td>Years of Data Collection</td>
</tr>
</tbody>
</table>

The important thing to keep in mind is that the ACS multiyear estimates describe the population and characteristics of an area for the full three- or five-year period, not for any specific day or period within the set of years, nor for any single year within the multi-year period.

### Differences with Single Year ACS Estimates

ACS single year estimates represent characteristics based on information collected over a single calendar year, whereas ACS multiyear estimates represent characteristics based on information collected over a three year or a five year period. The single year estimate will tend to provide more current information about areas that have changing population and housing characteristics, while multiyear estimates will provide more precise estimates for population and housing characteristics, and for more areas.

As the ACS multiyear estimates are based on data from a full 36 or 60 month time period, even if the ACS single year is within the ACS multiyear period (e.g., 2006 single year, 2006-2008 multiyear), it is unlikely that single year and multiyear estimates will be the same; this will be true even if the ACS single year is the mid-year of ACS multiyear period (e.g., 2007 single year, 2006-2008 multiyear).

For example, assume an area with a growing Hispanic population, with single year and three year estimates of percent of population who speak Spanish at home as provided in Table 2. The three-year estimates for a period differ from the estimates for the single years that comprise the period.
<table>
<thead>
<tr>
<th></th>
<th>Year Estimate</th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Year</td>
<td></td>
<td>13.1</td>
<td>13.6</td>
<td>13.7</td>
<td>15.1</td>
<td>15.9</td>
<td>16.8</td>
</tr>
<tr>
<td>Estimate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimate</td>
<td></td>
<td>13.4</td>
<td>14.1</td>
<td>14.9</td>
<td>15.9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 3. Differences Between ACS and Decennial Census Sample Data

Both the ACS and the Decennial Census sample data are based upon information from a sample of the population. The Decennial Census sample data were obtained from the census long form questionnaire, which was collected from about one in six households.

There are four major factors, discussed below, that lead to potential differences between the ACS and the Decennial Census. Two of the factors are related to definitional differences between the ACS and the decennial census—where people are considered to reside, and the reference periods for variables being collected. The other two factors have to do with the data collection time frame. The ACS generates estimates for characteristics of the population and housing based upon data collected over single and multiple calendar years. The ACS single year estimates contrast with the “point-in-time” estimates from the decennial census designed to measure characteristics of the population and housing from data collected around April 1 of the census year, reflecting a narrower timeframe than the ACS.

Census Bureau subject matter specialists have reviewed all these factors and have determined that, in general, ACS estimates are similar to those obtained from decennial census data for an area or characteristic. Thus, the practitioner should consider whether the particular analysis involves an area or characteristic that would be affected by these factors, or whether it would fit the pattern of most areas and characteristics which are not affected.¹

A. Residence Rules

Residence rules are used to determine which individuals are to be considered residents of a particular housing unit or group quarters. While most people have a definite and long-term tie to a housing unit or group quarters, there are many people who may stay in different places over the course of the year. For example, migrant workers move with crop seasons and do not dwell in any one location for the entire year.

The Decennial Census uses a “usual residence” rule – defined as the place where a person lives or stays most of the time as of Census Day – whereas the ACS uses a "two-month" residence rule – defined as anyone currently living or staying in the sample housing unit at the time of the interview and whose time in residence has or will exceed two months.

B. Reference Period

¹ Further information concerning areas and characteristics which do not fit the general pattern of comparability can be found on the ACS website (http://www.census.gov/acs/www/UseData/compACS.htm)
1. Income Data

To estimate annual income, the Census 2000 long form sample used the calendar year prior to Census Day as the reference period, whereas the ACS uses the 12 months prior to the interview date as the reference period. Thus, while the Census 2000 long form collected information for income relative to 1999, the single year ACS income estimates will contain a mixture of 12 month reference periods ranging from, in the case of the 2006 ACS single year estimates, full calendar year 2005 to December 2005 through November 2006. The ACS income estimates will thus reflect a melding across a longer time period than Census 2000.

2. Other Data

Census 2000 used the calendar year 1999 to estimate sales of agricultural products, while the ACS references sales during the 12 months prior to the interview date. The reference periods for two utility cost questions—gas and electricity—also differ between Census 2000 and the ACS, the census asking for annual costs while ACS asks for the cost last month. School attendance also differs (within last three months for ACS, since February 1, 2000 for Census 2000). The practitioner should consider the potential impact these reference period differences could have on distributions if comparing ACS estimates with Census 2000.

C. Definition for Data Items

There are some data items collected by both the ACS and the Census 2000 long form, but with slightly different definitions that could affect comparability of results, such as annual costs for a mobile home (Census 2000 included installment loan costs, ACS does not). These definitional differences could affect the level of the estimates for these items (in the case of annual costs for a mobile home, the ACS could be expected to yield smaller estimates than Census 2000).

D. Seasonal Variation in Characteristics

For population characteristics of an area that are affected by seasonal variation (e.g., age distribution, income, employment status) the Decennial Census will reflect the situation around Census Day, whereas the ACS will incorporate annual seasonal variability into the calendar year period estimate for the housing unit population meeting the 2 month rule. While most areas would not be affected by seasonal variation, small resort areas could evidence different distributions under the ACS.
Appendix 4. Sampling Errors

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 estimates that would have been obtained had a different set of sample units been selected for the survey.

There are various ways in which sampling error can be expressed quantitatively. As the ACS estimates are based upon 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 following information describing each of these sampling error measures, explaining how they differ, and explaining how each should be used is intended to assist the practitioner with analysis and interpretation of ACS estimates.

A. 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 “true value” in the population, as not all members of the population were measured in the survey. The standard error provides a quantitative measure of the extent to which an estimate derived from the sample survey can be expected to deviate from the “true value” in the population. The SE is not used for interpretation, but rather is the foundational measure from which the remaining sampling error measures are derived, and is used in exact tests of significance for making comparisons.

A very basic example of the standard example is a population with three units, with values of 1, 2, and 3; thus the average value for the 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 smaller areas are only published using multiple years of data, so as to take advantage of the larger sample size that results from aggregating data from more than one year.
SEs are used in exact tests of significance for making comparisons, and in the calculation of the coefficient of variation for an estimate as described in section D.

B. 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 "true value" for the population is within a certain distance (i.e., the MOE) of the sample estimate. Confidence levels of 90%, 95%, and 99% are commonly used in practice to lessen the risk associated with an incorrect inference. The MOE is important to protect against misinterpreting small or nonexistent differences as meaningful. It provides a concise measure of the precision for purposes of presentation with the sample estimate in a table, and is easily used to construct confidence intervals and make quick tests of significance.

For the ACS, the Census Bureau chose to provide the MOE at the 90 percent confidence level as the published sampling error measure. This can be roughly interpreted as meaning the user can be 90 percent certain that the uncertainty in the estimate due to sampling is equal to the value of the MOE.

The MOE is a multiple of the SE, with the multiplier depending upon the confidence level being used. The formula for the MOE at the 90 percent confidence level is

\[ MOE = 1.645 \times SE \]

For example, if the SE for an estimate was 1,000, then the MOE at the 90 percent confidence level would be 1,645 (=1.645 × 1,000).

The MOE is used in assessing the usability of an estimate, and in constructing a confidence interval for an estimate, as described in the next section.

C. A Confidence Interval (CI) defines a range expected to contain the “true value” for the population with a given level of confidence. A sample estimate and its MOE are used to construct the CI around the estimate. These intervals are ranges that are expected to contain the average value of the estimated characteristic that results over all possible samples, with a known probability. The CI is useful when graphing estimates, to show the extent of sampling error present in the estimates, and for visually comparing estimates.

The CI is constructed by adding to and subtracting from a sample estimate its corresponding MOE.

For example, if an estimate of 20,000 had an MOE at the 90 percent confidence level of 1,645, the CI would range from 18,355 (20,000 – 1,645) to 21,645 (20,000 + 1,645).
The confidence level associated with the MOE used in constructing the CI roughly describes the likelihood the “true value” for the population is contained within the CI.

For example, given a 90 percent confidence level was associated with the MOE used in constructing the CI above, the practitioner could roughly be 90 percent certain the “true value” of the population was between 18,355 and 21,645.

A CI is constructed to provide a range within which the population value can be expected to lie, with a certain likelihood, or when graphing estimates.

**D. A Coefficient of variation (CV)** provides an indication 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. It is a useful barometer of the stability, and thus the usability, of a sample estimate, and for deciding between use of single year and multiyear estimates.

In general, the smaller the CV, the more reliable the estimate is. A small CV indicates the sampling error is small relative to the estimate, and thus the user can be more confident that the estimate is close to the “true value” for the population. In many applications, a CV of 10 percent or less for an estimate is desirable – thus use a multiyear estimate will often be preferable to use of a single year estimate. In practice, however, there will be many estimates with CVs well over this desirable level, even with use of multiyear estimates. Although the practitioner should be cautious before using estimates with CVs much greater than 10 percent, a common minimum standard is to restrict use to estimates with CVs less than 50 percent.

For example, if an estimate of 20,000 had a SE of 1,000, then the CV for the estimate would be 5 percent ([1,000 / 20,000] x 100 percent).

CVs should be calculated to determine the usability of a sample estimate.

One cautionary note on the use of CV’s for small estimates: for estimates near zero (as can happen when looking at estimates of proportions), the CV will become very unstable and should not be used. The Census Bureau recommends that for estimates below 5 percent of the total population; the CV should not be used and that instead inferences should instead be based on the MOE or the SE.
Appendix 5. Deriving Sampling Error Measures

A. Deriving Margins of Error for Alternative Confidence Levels

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 “true value” for the population is within a certain distance (i.e., the MOE) of the sample estimate.

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 the 90 percent confidence level. However, practitioners may want to use other confidence levels, such as 95 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 likelihood of a 10 percent chance of an incorrect inference, versus a 1 percent chance if using a 99 percent confidence level. Thus, if the impact of an incorrect conclusion is large, the practitioner should consider increasing the confidence level used.

One commonly experienced situation where use of a 95 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 incorrect inferences.

If a MOE corresponding to a confidence level other than 90 percent is desired, the published MOE must be converted to the appropriate MOE corresponding to the desired confidence level. This is accomplished by multiplying the published MOE by an adjustment factor. If the desired confidence level is 95 percent, then the factor is equal to 1.96 / 1.645; if the desired confidence level is 99 percent, then the factor is equal to 2.58 / 1.645. (NOTE: If working with ACS single year estimates for 2005 or earlier, use the value 1.65 rather than 1.645 in the adjustment factor².)

Conversion of MOE’s can be expressed as

\[
MOE_{0.95} = \frac{1.96}{1.645} \times MOE_{ACS} \\
MOE_{0.99} = \frac{2.58}{1.645} \times MOE_{ACS}
\]

² The value 1.65 must be used for ACS single year estimates for 2005 or earlier as that was the value used in deriving the published MOE from the standard error in those years.
where $MOE_{ACS}$ is the ACS published MOE for the estimate

For example, the ACS published MOE for estimated number of civilian veterans in the state of Virginia from the 2006 ACS is 12,357. The MOE corresponding to a 95 percent confidence level would be derived as

$$MOE_{95} = \frac{1.96}{1.645} \times 12,357 = 14,723$$

B. Deriving the Standard Error from the MOE

The standard error (SE), which measures the variability of an estimate due to sampling, is the foundational measure from which the remaining sampling error measures are derived. As such, the SE is used when comparing estimates to determine whether the differences between the estimates can be said to be statistically significant, or when assessing the precision of an estimate.

When conducting exact tests of significance or to calculate the CV for an estimate (see Section D), the SE’s of the estimates are needed. To derive the SE, simply divide the published MOE by 1.645. (NOTE: If working with ACS single year estimates for 2005 or earlier, use the value 1.65 rather than 1.645 in the adjustment factor-refer to footnote 1.)

Derivation of SE’s can thus be expressed as

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

where $MOE_{ACS}$ is the ACS published MOE for the estimate

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

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

C. Constructing Confidence Intervals

A confidence interval (CI) defines a range expected to contain the “true value” for the population with a given level of confidence. CIs are constructed by adding to and subtracting from a sample estimate its corresponding MOE. Thus, to construct a CI at the 90 percent confidence level, the lower bound is defined as the estimate minus the published MOE, and the upper bound is defined as the estimate plus the published MOE. For CIs at the 95 or 99 percent confidence level, the appropriate MOE must first be derived as explained above.
Construction of the lower and upper bounds for the CI can thus be expressed as

\[ L_{CL} = \hat{X} - MOE_{CL} \]
\[ U_{CL} = \hat{X} + MOE_{CL} \]

where \( \hat{X} \) is the ACS estimate
\( MOE_{CL} \) is the MOE for the estimate at the desired confidence level

For example, to construct a CI at the 95 percent confidence level for the number of civilian veterans in the state of Virginia in 2006, one would use the 2006 estimate (771,782), and the corresponding MOE at the 95 percent confidence level derived above (14,723)

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

### D. Calculating Coefficients of Variation

The coefficient of variation (CV) for an estimate provides an indication of the relative amount of sampling error that is associated with a sample estimate. The CV is calculated by dividing the SE for an estimate by the estimate itself, and then multiplying by 100%. The method for obtaining the SE for an estimate was described above.

Thus, the CV can be expressed as

\[ CV = \frac{SE}{\hat{X}} \times 100\% \]

where \( \hat{X} \) is the ACS estimate
\( SE \) 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 above (7.512)

\[ CV = \frac{7.512}{771,782} \times 100\% = 0.1\% \]

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

Information on significance testing is provided in Appendix 8.
Appendix 6. Guidance on Use of Single and Multiyear ACS Estimates

Two primary uses of ACS estimates are for assessing change over time and for understanding characteristics of the population in order to make local decisions such as the locations of schools, or hospitals, to understand emerging populations or potential markets to determine need for services or new businesses, to carry out transportation analysis, etc. In the past, decennial census sample data provided the most comprehensive information. However, the currency of those data suffered through the inter-censal period and the ability to assess change over time was limited. The availability of the ACS estimates greatly improves the currency of data for understanding characteristics of the population, and enhances the ability to assess change over time.

For practitioners trying to decide whether to use single or multiyear ACS estimates for areas where both are available, there are several key factors: intended use of the estimates; precision of the estimates; and currency of the estimates. All of these factors, along with an understanding of the key differences between single year and multiyear ACS estimates, should be taken into consideration when deciding which time period to use.

For practitioners interested in estimates for smaller geographic areas, the multiyear ACS estimates will be the only option for consideration. For the very smallest areas, the only option will be to use the five-year ACS estimates, while for small areas with populations of at least 20,000, the three-year ACS estimates will also be an option.

A. Key Differences Between Single Year and Multiyear Estimates

The key trade-off to be made in deciding whether to use single year or multiyear estimates is between currency and precision. While single year estimates offer more current estimates, they will also have higher coefficient of variations (CVs) roughly 70 percent larger than those for three-year estimates. For single year estimates with small CVs (less than 10 percent), this should not be a critical factor. In general, the single year estimates are preferred as they will be more relevant to the current conditions.

However, the practitioner must take into account the level of uncertainty present in the single year estimates, which may be great for small subpopulation groups. A single year estimate with a CV\(^3\) less than 10 percent is usually preferable to a multiyear estimate as it is more up to date. However, multiyear estimates should be used where possible when the single year estimate has a CV greater than 10 percent. If the CV for the single year estimate is greater than 50 percent, then the single year estimate should not be used at all.

\(^3\) The CV is calculated as [standard error / estimate] x 100%
Table 1 illustrates how the most current five-year MYE can differ from the most current single-year estimate. It shows the percentage of households where Spanish is spoken at home for test counties Broward, FL, and Lake, IL. The differences between the 2004 and 2000-2004 estimates are not small at 1.4 and 1.5 percentage points.

<table>
<thead>
<tr>
<th>County</th>
<th>2004 SE of 2004</th>
<th>2000-2004 Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Broward</td>
<td>19.9 0.2</td>
<td>18.5 1.4</td>
</tr>
<tr>
<td>Lake</td>
<td>15.9 0.2</td>
<td>14.4 1.5</td>
</tr>
</tbody>
</table>

In this illustration, the CVs for the single year estimates (1.0 percent (=0.2/19.9) for Broward and 1.3 percent (=0.2/15.8) for Lake) are sufficiently small to warrant 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 large CV’s. In general, multiyear estimates are preferable to single year estimates when looking at estimates based on 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 cutoff for publishing single-year estimates for geographic areas. However, some subpopulations will be much smaller than 65,000. In Table 2 we see that there are an estimated 21,881 families based on the 2000-2004 multiyear estimate; but the number of families with a female householder, no husband, with related children less than 18 years, has an estimate of only 1,883. Not surprisingly, the 2004 single-year estimate of the poverty rate (38.3 percent) for this subpopulation has a large standard error (SE), 13.0 percentage points, and thus the CV is 33.9 percent (=38.3/13.0), much greater than the 10 percent CV guideline for use.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>All families</td>
<td>21,881</td>
<td>9.5 0.8</td>
<td>10.0 2.3</td>
</tr>
<tr>
<td>With related children under 18 years</td>
<td>9,067</td>
<td>15.3 1.5</td>
<td>17.8 4.5</td>
</tr>
<tr>
<td>Married couple families</td>
<td>17,320</td>
<td>5.6 0.7</td>
<td>7.9 2.0</td>
</tr>
<tr>
<td>With related children under 18 years</td>
<td>6,633</td>
<td>7.7 1.2</td>
<td>12.1 3.9</td>
</tr>
<tr>
<td>Families with female householder, no husband</td>
<td>3,433</td>
<td>27.2 3.0</td>
<td>19.0 7.2</td>
</tr>
<tr>
<td>With related children under 18 years</td>
<td>1,883</td>
<td>40.2 4.9</td>
<td>38.3 13.0</td>
</tr>
</tbody>
</table>

For such small subpopulations users obtain much more precision using the three-year or five-year estimate. In this example the five-year estimate of 40.2 percent has a SE of 4.9 percentage points (which yields a CV of 12.2 percent (=40.2/4.9)), and the three-year estimate of 40.4 percent has a SE of 6.8 percentage points (which yields a CV of 16.8 percent (=40.4/6.8)).
Multiyear estimates should not be used for estimating the year to year change in a particular characteristic. This is because roughly two-thirds of the data in a three-year estimate from one year overlaps with the data in the next year’s three year estimate (the overlap is roughly four-fifths for five-year estimates). Thus, differences in overlapping multiyear estimates are driven by differences in the non-overlapping years (e.g., 2006 and 2009 when comparing 2006-2008 and 2007-2009 three-year estimates) rather than by differences in the successive calendar years, as illustrated in Figure 1 (in this example the difference of interest would be between 2008 and 2009; however, the difference in the multiyear estimates would relate to 2006 and 2009).

**Figure 1**: Data Collection Periods for Three-Year Estimates

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-2008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007-2009</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Variability in single year estimates for smaller areas (near the 65,000 publication threshold) and small subgroups within even large areas may limit the ability to examine trends. For example, single year estimates for a characteristic with a CV well above 10 percent may bounce up and down, making it difficult to see the 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. The smoothing effect of multiyear estimates is illustrated in Figure 2.
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.

**B. Further Notes on Considering the Reliability of Estimates**

As mentioned in Section A, a CV of 10 percent or less for an estimate is desirable. In practice, however, there will be many estimates with CVs well over this desirable level. Although the practitioner should be cautious before using estimates with CVs much greater than 10 percent, a common minimum standard is to restrict use to estimates with CVs less than 50 percent.

A general guideline when working with five-year ACS estimates is that individual block- and tract-level estimates should not be used as they will tend to be unstable even with five years' worth of sample data. Rather, the practitioner should aggregate such estimates to a neighborhood level to improve the reliability of the estimates.

Practitioners should examine the CV associated with an estimate prior to using the estimate, using the paragraph above as a “fitness for use” guideline. The critical issue associated with an estimate having a CV greater than 50 percent is that at the 95 percent confidence level its confidence interval would contain zero, indicating the estimate is extremely unstable.
C. Guidelines Concerning the Use of Single Year Estimates

Single year estimates should, in general, be used for larger geographies and populations, and for examining year to year changes for estimates with small CV’s. Given the availability of a single year estimate, one should calculate the CV and determine if it or the MOE appears so large that the single year estimate should not be used. In general, single year estimates should not be used for smaller subpopulations in larger geographies.

Given the sampling rates used for ACS, as a rough rule of thumb the practitioner would be cautioned against working with data for a subpopulation with a total estimated size less than 3,000 for single year estimates, 1,000 for three-year estimates, and 600 for five-year estimates, as the sample size upon which the estimates are based may be less than 50 individuals.

While a practitioner may determine it is appropriate to use a single year estimate for an analysis, based upon the intended use and CV associated with the estimate as discussed earlier in this chapter, it is important 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.

D. Guidelines Concerning the Use of Multiyear Estimates

Multiyear estimates should, in general, be used when single year estimates have large CVs, for smaller geographies and smaller populations in larger geographies, for examining change over non-overlapping time periods, and for smoothing data trends over time.

Multiyear estimates in general should not be used to examine year to year changes, nor to examine any other overlapping time periods. 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.

E. Considerations Involving Local Area Variations

Local area variations may occur due to rapidly occurring changes. As discussed previously, multiyear estimates will tend to be insensitive to such changes when they first occur. Single year estimates, if associated with sufficiently small CVs, can be very valuable in identifying and studying such phenomena. Graphing trends for such areas using single year, three-year, and five-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 single year estimates will eventually be incorporated into 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.
Appendix 7. Calculating Margins of Error for Derived Estimates

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

ACS tabulations provided through AFF contain the associated margins of error (MOEs) at the 90 percent confidence level. However, when derived estimates are generated (e.g., aggregate estimates, or proportions, or ratios not available within AFF), the user must calculate the MOE for the derived estimate.

The 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 “true value” for the population is within a certain distance (i.e., the MOE) of the sample estimate. The MOE is important to protect against misinterpreting small or nonexistent differences as meaningful. MOEs for derived aggregate counts, proportions, and ratios can be easily calculated using the formulae that follow.

Calculated MOEs based on the MOEs provided in AFF for the components of the derived estimates will be at the 90 percent confidence level. If a 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 a MOE for the desired confidence level as described in Appendix 5.

A. Calculating MOEs for Aggregated Count Data

To calculate the MOE for aggregated count data, do the following:

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:

\[ \text{MOE}_{agg} = \sqrt{\sum_c MOE_c^2} \]

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>
<thead>
<tr>
<th>Characteristic</th>
<th>Estimate</th>
<th>MOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Females living alone in Fairfax County (Component 1)</td>
<td>52,354</td>
<td>3,303</td>
</tr>
<tr>
<td>Females living alone in Arlington County (Component 2)</td>
<td>19,464</td>
<td>2,011</td>
</tr>
<tr>
<td>Females living alone in Alexandria City (Component 3)</td>
<td>17,190</td>
<td>1,854</td>
</tr>
</tbody>
</table>

The aggregate estimate is:

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

Obtain MOEs of the component estimates:

\[
\text{MOE}_{\text{Fairfax}} = 3,303, \quad \text{MOE}_{\text{Arlington}} = 2,011, \quad \text{MOE}_{\text{Alexandria}} = 1,854
\]

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

\[
\text{MOE}_{\text{agg}} = \sqrt{(3,303)^2 + (2,011)^2 + (1,854)^2} = \sqrt{18,391,246} = 4,289
\]

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

**B. Calculating MOEs for Derived Proportions and Ratios**

The difference between a proportion and a ratio is that in a proportion, the numerator is a subset of the denominator (e.g., the proportion of single person households that are female), whereas in a ratio, the numerator is independent of the denominator (e.g., the ratio of females living alone to males living alone).

**1. Derived Proportions**

To calculate the MOE for derived proportions, do the following:

1) Obtain the MOE for the numerator and 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 from 5) from the squared MOE of the numerator
7) Take the square root of the result from 6)
8) Divide by the numerator of the proportion

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

\[
MOE_p = \sqrt{\frac{MOE_{num}^2 - \hat{p}^2 \cdot MOE_{den}^2}{\hat{X}_{num}}}
\]

where \( MOE_{num} \) is the MOE of the numerator
\( MOE_{den} \) is the MOE of the denominator
\( \hat{p} \) is the derived proportion
\( \hat{X}_{num} \) is the estimate used in the numerator of the derived proportion
\( \hat{X}_{den} \) is the estimate used in the denominator of the derived proportion

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

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Estimate</th>
<th>MOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black females 25 years and over with a graduate degree (Numerator)</td>
<td>4,634</td>
<td>989</td>
</tr>
<tr>
<td>Black females 25 years and over (Denominator)</td>
<td>31,713</td>
<td>601</td>
</tr>
</tbody>
</table>

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 and over in Fairfax County with a graduate degree
\( \hat{X}_{BF} \) is the ACS estimate of black females 25 years and over in Fairfax County

Obtain MOE's of the numerator (number of black females 25 years and over in Fairfax County with a graduate degree) and denominator (number of black females 25 years and over in Fairfax County)

\[ MOE_{num} = 989, \ MOE_{den} = 601 \]

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

\[ MOE_{num}^2 - \hat{p}^2 \cdot MOE_{den}^2 = (989)^2 - (0.1461)^2 \cdot (601)^2 = 978,121 - 7712.3 = 970,408.7 \]
Calculate the MOE by dividing the square root of the prior result by the numerator
\[
MOE_p = \frac{\sqrt{970,408.7}}{4,634} = \frac{985.1}{4,634} = 0.2126
\]

Thus, the derived estimate of the proportion of black females with a graduate degree in Fairfax County, Virginia is 0.1461, and the MOE for the estimate is 0.2126. As explained in Appendix 6, use of multiyear estimates would be recommended in this example as the MOE is greater than 50 percent the size of the estimate.

2. Derived Ratios

To calculate the MOE for derived ratios, do the following:

1) Obtain the MOE for the numerator and 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 from 5) to the squared MOE of the numerator
7) Take the square root of the result from 6)
8) Divide by the numerator of the ratio

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

\[
MOE_R = \frac{\sqrt{MOE_{num}^2 + \hat{R}^2 * MOE_{den}^2}}{\hat{X}_{num}}
\]

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 in the numerator of the derived ratio
\( \hat{X}_{den} \) is the estimate used in the denominator of the derived ratio

The example below shows how to derive the MOE for the estimated ratio of black females 25 years and over Fairfax County, Virginia with a graduate degree to black males 25 years and over in Fairfax County with a graduate degree, based on the 2006 ACS
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Estimate</th>
<th>MOE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black females 25 years and over with a graduate degree (Numerator)</td>
<td>4,634</td>
<td>989</td>
</tr>
<tr>
<td>Black males 25 years and over with a graduate degree (Denominator)</td>
<td>6,440</td>
<td>1,328</td>
</tr>
</tbody>
</table>

The estimated ratio is:

\[
\hat{p} = \frac{\hat{X}_{\text{gradBF}}}{\hat{X}_{\text{gradBM}}} = \frac{4,634}{6,440} = 0.7200
\]

Obtain MOE's of the numerator (number of black females 25 years and over with a graduate degree in Fairfax County) and denominator (number of black males 25 years and over in Fairfax County with a graduate degree)

\[MOE_{\text{num}} = 989, \quad MOE_{\text{den}} = 1,328\]

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

\[
MOE_{\text{num}}^2 - \hat{R}^2 \cdot MOE_{\text{den}}^2 = (989)^2 + (0.7200)^2 \cdot (1,328)^2 = 978,121 + 472,799.3 = 1,450,920.3
\]

Calculate the MOE by dividing the square root of the prior result by the numerator

\[
MOE_p = \frac{\sqrt{1,450,920.3}}{4,634} = \frac{1,204.5}{4,634} = 0.2600
\]

Thus, the derived estimate of the ratio of number of black females 25 years and over in Fairfax County, Virginia with a graduate degree to the number of black males 25 years and over in Fairfax County, Virginia with a graduate degree is 0.7200, and the MOE for the estimate is 0.2600.
Appendix 8. Making Comparisons

One of the most important uses of the ACS estimates is to make comparisons between estimates. There are several key types of comparisons that are of general interest to practitioners: 1) comparisons of estimates from different geographic areas within the same time period (e.g., comparing proportion of persons below the poverty level in two counties); 2) comparisons of estimates for the same geographic area across time periods (e.g., comparing proportion of persons 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 proportion of persons below the poverty level in a county for 2006 and 2000).

There are a number of conditions which 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 differences are statistically significant, rather than just the looking at the differences in the estimates. Statistical significance means that there is statistical evidence a true difference exists, and that the observed difference is unlikely to have occurred by chance. Other conditions depend upon the types of comparisons being made.

A method for determining statistical significance (i.e., determining whether the difference between two estimates can be said to represent a difference that exists within the full population or whether the difference occurred by chance due to sampling) when making comparisons is presented in the next section. Discussion of considerations associated with the various types of comparisons which could made then follows.

A. Determining Statistical Significance

When comparing two estimates, one should use the test for significance described below. This approach will allow the practitioner to ascertain whether the observed difference is likely due to chance (and thus is not statistically significant) or most 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 to carry out using data provided within tables generated from American FactFinder.

1) Determine the SE for each estimate (for ACS, SE is defined by margin of error (MOE)/1.6454)

\[^4\text{NOTE: If working with ACS single year estimates for 2005 or earlier, use the value 1.65 rather than 1.645.}\]
2) Square the resulting SE for each estimate
3) Sum the squared SEs
4) Take square root of sum of squared SEs
5) Subtract estimate 2 from estimate 1
6) Divide difference between estimates by square root of sum of squared SEs
7) Compare the absolute value of the result from 6) with the critical value for the desired level of confidence (1.645 for 90 percent, 1.96 for 95 percent, 2.58 for 99 percent)
8) If the absolute value of the result from 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:

\[
\frac{\hat{X}_1 - \hat{X}_2}{\sqrt{SE_1^2 + SE_2^2}} > Z_{cl},
\]

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

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

\( Z_{cl} \) is the critical value for the desired confidence level (≈1.645 for 90 percent, 1.96 for 95 percent, 2.58 for 99 percent)

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

\[
\frac{22.0 - 21.5}{\sqrt{(0.12)^2 + (0.12)^2}} = \frac{0.5}{\sqrt{0.015 + 0.015}} = \frac{0.5}{\sqrt{0.03}} = \frac{0.5}{0.173} = 2.90
\]

As the test value (2.90) is greater than the critical value for a confidence level of 99 percent (2.58), then the difference in percentages is statistically significant at a 99 percent confidence level (also referred to as statistically significant at the alpha=0.01 level). A rough interpretation of the result is that the practitioner can be 99 percent certain that a difference exists between the proportion of households with one or more people of age 65+ 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
As the test value (1.25) is less than the critical value for a confidence level of 90 percent (1.645), then the difference in percentages is not statistically significant. A rough interpretation of the result is that the practitioner cannot be certain to any sufficient degree that the observed difference in the estimates was not due to chance.

**B. Comparisons Within the Same Time Period**

Comparisons involving two estimates from the same time period (e.g., from the same year or same three-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 practitioners may want to be aware of.

When comparing estimates within the same time period, the areas or groups will generally be non-overlapping (e.g., comparing estimates for two different counties, or for two different age groups). In these cases, 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, comparing an estimate for females in total 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 practitioner has reason to believe the two estimates are strongly correlated, he/she is advised to ignore the partial dependence and use the formula for testing differences as provided in the previous section. If the two estimates are believed to be strongly correlated, a finding of statistical significance will not be affected by the correlation, whereas a finding of lack of statistical significance based on the formula may be incorrect. If the practitioner feels it important to obtain a more exact test of significance, he or she should consult with a statistician as to approaches for accounting for the correlation in performing the statistical test of significance.

**C. Comparisons Across Time Periods**

Comparisons across time periods may involve two single year periods or two of the same multiyear periods within the same area. Comparisons across time periods should be made only with comparable period estimates. When carrying out any of these types of comparisons across time periods, there are several
issues that practitioners should be aware of and take into consideration. Practitioners 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 are difficult to interpret.

When comparing estimates from two different single year periods, if one of the years is prior to 2006 and the other is 2006 or after (e.g., comparing estimates from 2005 and 2007), the practitioner should recognize that beginning with 2006 the ACS includes the population living in group quarters (GQ) as well as the population living in housing units. A GQ 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, and residency is commonly restricted to those receiving these services. People living in group quarters are usually not related to each other. Group quarters include such places as college residence halls, residential treatment centers, skilled nursing facilities, group homes, military barracks, correctional facilities, workers’ dormitories, and facilities for people experiencing homelessness. Many types of GQ populations have demographic, social, or economic characteristics that are very different from the household population.

The inclusion of the GQ population in the 2006 ACS could therefore have a noticeable impact on the distributions of these characteristics. This is particularly true for areas with a substantial GQ population. For most population subjects, the Census Bureau suggests that you make comparisons only if the geographic area of interest does not include a substantial GQ population. For most housing subjects, the Census Bureau supports comparisons being made.

The primary consideration to be made when comparing estimates from two multiyear periods is that practitioners should limit their comparisons to non-overlapping periods (e.g., comparing estimates from 2006-2008 with estimates from 2009-2011). The reason for this recommendation is that with overlapping periods the interpretation of differences is much more complex than for non-overlapping periods (as part of the data is used in the estimates for each time period), and the test of significance formula should account for the lack of independence between the two estimates.

D. Comparisons with Census 2000 Data

In Appendix 3, major differences between ACS and Decennial Census sample data were discussed. Such factors as differences in residence rules, universes, and reference periods, while not discussed in detail in this Appendix, should be considered when making comparisons of ACS estimates to Decennial Census data. Given reference period differences, there is also the potential for
seasonality affecting comparisons between Decennial Census ACS estimates when looking at smaller resort areas, etc.

However, Census Bureau subject matter specialists have reviewed all the factors affecting differences between ACS and Decennial Census estimates and have determined that, in general, ACS estimates would be similar to those obtained from Decennial Census data for an area or characteristic. Thus, the practitioner should consider whether the particular analysis involves an area or characteristic that would be affected by these differences, or whether it would fit the pattern of most areas and characteristics which are not affected.5

When comparing ACS and Decennial Census sample data, the practitioner must remember that the Decennial Census sample data have associated standard errors as do the ACS estimates, and must incorporate the standard errors for both estimates in performing tests of statistical significance. Section A of this Appendix provides the calculations necessary for determining statistical significance of a difference between two estimates. In the case of the Decennial Census sample data estimate, the SE would be derived using 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 driving 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.

E. Comparisons with Census 2010 Data

Looking ahead to the 2010 Decennial Census, a major factor for consideration is that long form variables will no longer be collected in the 2010 Decennial Census, and thus will not be available for comparison with ACS estimates. The only common variables for the ACS and 2010 Decennial Census are sex, age, race, ethnicity, relationship, and housing tenure.

The critical factor that must be considered when comparing ACS estimates encompassing 2010 with the 2010 Decennial Census is the potential impact of the population controls used for ACS. As the population controls used for 2010 ACS data will be based off the population estimates program, they will not agree

---

5 Further information concerning areas and characteristics which do not fit the general pattern of comparability can be found on the ACS website (http://www.census.gov/acs/www/UseData/compACS.htm)
with the 2010 Decennial Census population counts for that year. 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 Decennial Census population counts is greater.

F. Comparisons with Other Surveys

Comparisons of ACS estimates with estimates from other national surveys, such as the Current Population Survey, may be of interest for some practitioners. A major consideration in making such comparisons will be that ACS estimates include data for populations in both institutional and noninstitutional group quarters, whereas 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 Section A. The standard error for the estimate from the other survey would be derived according to technical documentation provided for that survey.

Finally, the practitioner wishing to compare ACS estimates with estimates from other national surveys should consider the potential impact that other factors, such as target population, sample design and size, survey period, reference period, residence rules, and interview modes, could have on estimates from the two sources.

G. A Note on Multiple Comparisons

When comparisons are involved in analysis, it is often the situation that many comparisons are being made in the same data set (e.g., comparing differences by age group for several race and ethnicity categories). Nor is it uncommon for a practitioner to review a set of tabulations, looking for significant differences. However, the confidence level associated with a comparison assumes that only one comparison has been identified for analysis. As a result, the actual confidence level when multiple comparisons are made will be less than the confidence level used for the test.

While exact methods exist for adjusting the critical value for significance testing, as a rough rule of thumb, larger confidence level should be used for comparisons. Preferably, a confidence level of 95 or 99 percent should be used unless the practitioner is truly only interested in making one comparison. This is because, if a 90 percent confidence level is used, one would expect that roughly 10 percent of differences identified as statistically significant would represent true
differences, and use of larger confidence levels would reduce the potential for false readings.

As an illustration, suppose the proportion of persons 65+ in County Z was compared to the proportion for each of the other 60 counties in the state, using a significance level of 90 percent. The findings may indicate that the proportion for County Z was significantly different from the proportion of 8 other counties. However, using a significance level of 90 percent implies there is a 10 percent chance of a false finding. Thus, one could expect a significant difference finding for 6 of the 60 counties just by chance. If instead, a significance level of 99 percent were used in carrying out each of the comparisons, the expectation of false findings would be for less than one county, allowing the practitioner to place more confidence in the findings of significant differences.
Appendix 9. Using Dollar-denominated Data

Dollar-denominated data refer to any characteristics for which inflation adjustments are used when producing or comparing estimates across time periods. For example, income, rent, value, and energy costs across time periods 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) to create estimates that reflect the dollar values at the end of the final year of the time period.

Given the potential impact of inflation on observed differences across time periods, when examining differences in estimates for dollar-denominated data from two time periods the practitioner should adjust for the effects of inflation on the observed differences. Such an adjustment will provide comparable estimates accounting for inflation. In making the adjustment, the Census Bureau recommends using the December All Items CPI-U-RS adjustment factors published annually by the Bureau of Labor Statistics (the link to BLS CPI adjustment factors through 2006 is: http://www.bls.gov/cpi/cpiurs1978_2006.pdf).

Adjustment of Single Year Estimates

For single year estimates from two different years, adjustment would be made as follows:

1) Obtain the December All Items CPI-U-RS for the two years being compared
2) Calculate the inflation adjustment factor as the ratio of most recent year CPI-U-RS from the most recent year to the CPI-U-RS from the least recent year
3) Multiply the dollar-denominated data estimated for the least recent year by the inflation adjustment factor

The inflation adjusted estimate for the least recent year can be expressed as:

\[
\hat{X}_{Y_1,\text{adj}} = \frac{CPI_{Y_2}}{CPI_{Y_1}} \hat{X}_{Y_1}
\]

where \(CPI_{Y_1}\) is the December All Items CPI-U-RS for the least recent year (Y1)
\(CPI_{Y_2}\) is the December All Items CPI-U-RS for the most recent year (Y2)
\(\hat{X}_{Y_1}\) is the published ACS estimate for the least recent year (Y1)

The example below compares the national median value for owner-occupied mobile homes in 2005 ($37,700) and 2006 ($41,000). One would first adjust the
2005 median value using the December 2005 All Items CPI-U-RS (196.8) and the December 2006 All Items CPI-U-RS (201.8), as follows:

\[
\hat{X}_{2005,\text{Adj}} = \frac{201.8}{196.8} \times 37,700 = 38,658
\]

Thus, the comparison of the national median value for owner-occupied mobile homes in 2005 versus 2006, in 2006 dollars, would be $38,658 (2005 inflation-adjusted) versus $41,000 (2006).

**Adjustment of Multiyear Estimates**

For multiyear estimates from two different time periods, adjustment would be made as follows:

1) Obtain the December All Items CPI-U-RS for the last year in each time period being compared
2) Calculate the inflation adjustment factor as the ratio of the CPI-U-RS from the most recent year to the CPI-U-RS from the least recent year
3) Multiply the dollar-denominated data estimate for least recent time period by the inflation adjustment factor

The inflation adjusted estimate for the least recent year can be expressed as:

\[
\hat{X}_{p_1,\text{Adj}} = \frac{CPI_{p_2}}{CPI_{p_1}} \times \hat{X}_{p_1}
\]

where \(CPI_{p_1}\) is the December CPI-U-RS for the last year in the least recent time period (P1)
\(CPI_{p_2}\) is the December CPI-U-RS for the last year in the most recent time period (P2)
\(\hat{X}_{p_1}\) is the published ACS estimate for the least recent time period (P1)

As an illustration, assume ACS multiyear estimates were available for 2001-2003 and 2004-2006. To compare the national median value for owner-occupied mobile homes in 2001-2003 (assume = $32,000) and 2004-2006 (assume = $39,000), one would first adjust the 2001-2003 median value using the December 2003 All Items CPI-U-RS (184.3) and the December 2006 All Items CPI-U-RS (201.8), as follows:

\[
\hat{X}_{2001-2003,\text{Adj}} = \frac{201.8}{184.3} \times 32,000 = 35,039
\]

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, as the inflation adjustment uses the All Items CPI, it will not reflect differences that may exist across characteristics such as energy and housing costs.

Another consideration in working with dollar-denominated data is the fact that amounts are adjusted using inflation factors based on the Consumer Price Index (CPI) to create estimates that reflect the dollar values at the end of the final year of the time period. To the extent that differences from the CPI-U-RS exist for an area or an item, the inflation adjustment within a time period will affect the ACS estimates. The likelihood of such an effect will be greater for five-year estimates.
Appendix 10. Quality Measures

Survey estimates are subject to both sampling and non-sampling errors. In Appendix 4, the topic of sampling error and the various measures available for understanding the uncertainty in the estimates due to their being based on estimates derived from a sample, rather than from an entire population, were discussed. Other errors, referred to collectively as non-sampling errors, may occur in the course of collecting and processing the ACS that affect the overall accuracy of the sample estimates.

Broadly speaking, non-sampling error refers to any error affecting a survey estimate outside of sampling error, which is the error due to collecting data from a sample rather than from the full population. Non-sampling error can occur in complete censuses as well as in sample surveys. Non-sampling errors are commonly categorized as coverage error, unit nonresponse error, item nonresponse error, response error, and processing error.

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

The final ACS estimates are assumed to be adjusted for under- and over-coverage by controlling specific estimates to independent total housing unit controls and population controls by sex, age, race, and Hispanic origin (more information is provided on the coverage error definition page of the “ACS Quality Measures” website 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 pre-controlled 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 under-covered or over-covered housing units or individuals differ from those eligible to be selected, the ACS may not provide an accurate picture of the population prior to the coverage adjustment, and the population controls may thus not eliminate or minimize that coverage error.

2) Unit nonresponse is the failure to obtain the minimum required information from a housing unit or a group quarters in order for it to be considered a completed interview. Unit nonresponse means that no survey data are available for a particular sampled unit. For example, if no one in a sampled housing unit is available during the timeframe for data collection, 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.

3) **Item nonresponse** occurs when a respondent fails to provide an answer to a required item, or when the answer given is inconsistent with other information and thus unusable. Item nonresponse will also occur when no data are provided for a person within a sampled unit. 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 lack of understanding of a particular question by a respondent.

Information on item nonresponse allows practitioners 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 differ from those who do. The ACS estimation methodology includes imputations for item nonresponse, intended to reduce the potential for item nonresponse bias.

4) **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 for 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.

5) **Processing error** can occur during the preparation of the final data files. For example, errors may occur if data entry of information on the questionnaire must occur to create the survey data file. Coding of responses can also result in processing error.

Non-sampling errors can result in either random errors or systematic errors. Of greatest concern are systematic errors due to non-sampling error, as random errors are generally cancelled out at higher geographic levels if a large enough sample is used, as is the case with ACS.
Systematic errors are those errors that tend to accumulate over the entire sample. For example, if there is an error in the questionnaire design, this could cause problems with the respondent's answers, which in turn, can create processing errors, etc. Systematic errors often lead to a bias in the final results. Unlike sampling error and random error resulting from non-sampling error, bias caused by systematic errors cannot be reduced by increasing the sample size.

**ACS Quality Measures**

Non-sampling errors are extremely difficult, if not impossible, to measure directly. However, the Census Bureau has developed a number of indirect measures of non-sampling error to help inform practitioners of the quality of the ACS estimates: sample size; coverage rates; unit response rates and nonresponse rates by reason; and item allocation rates. All the measures are available on the “ACS Quality Measures” website (http://www.census.gov/acs/www/UseData/sse/).

**Sample size** measures for the ACS summarize information for the housing unit samples. The measures available at the state level are:

- **Housing Units**
  - Number of initial addresses selected
  - Number of final survey interviews

- **Group Quarters People** (beginning with the 2006 ACS)
  - Number of initial sample selected
  - Number of final survey interviews

In addition to the state sample size measures for both housing units and group quarters, county and county equivalent sample size measures for housing units are available on the “About the ACS” website (http://www.census.gov/acs/www/SBasics/); select link labeled “Sample Size” in the box on the left side of the frame).

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 4, there may be situations where the CV is small, but the practitioner has reason to believe the sample size for a subgroup could be very small.

---

6 The sample size measures for housing units (number of initial addresses selected and number of final survey interviews) cannot be used to calculate a response rate. The reason is that the number of initial addresses selected includes address that were determined to not identify actual housing units, and subsequent subsampling of initial addresses selected was sometimes carried out for CAPI nonresponse follow-up.
For example, Asians make up roughly 1 percent (8,418/656,700) of the population in Jefferson County, Alabama. Given the completed housing units in Jefferson county for the 2006 ACS were 4,972 and assuming roughly 2.5 persons per housing unit (or roughly 12,500 completed person interviews), one could estimate the 2006 ACS data for Asians in Jefferson county were based on roughly 150 completed person interviews.

**Coverage rates** are available by gender at both the state and national level, and for six race/ethnicity categories at the national level.

Low coverage rates provide an indication of the extent of adjustment to the weights for the sample data that were required during the population controls component of the estimation methodology, and thus indicate the greater potential for the presence of coverage error in the estimates. If characteristics of under-covered or over-covered housing units or individuals differ from those eligible to be selected, the ACS sample may not provide an accurate picture of the full population, and the population control adjustments may thus not eliminate or minimize coverage error.

**Unit response and nonresponse rates** are available at the state and national level by reason for refusal (refusal, unable to locate, no one home, temporarily absent, language problem, other, and insufficient data from an interview to be included in the data set). Rates are provided separately for housing units and for persons in group quarters.

A low unit response rate is an indicator of the potential for the presence of bias within the survey estimates for the state. For example, although the 2006 housing unit response rates are above 90 percent for all states, response rates for DC (91 percent) are lower than those for other states (94 percent and higher).

**Item allocation rates**, which roughly correspond to item nonresponse rates, are available at the state and national level for all individual questionnaire items, both housing and population. While item allocation rates are generally less than 10 percent at the national level, there are some exceptions, such as wage/salary income (13 percent in 2006).

Item allocation rates do vary by state, so practitioners are advised to examine the allocation rates for characteristics of interest before drawing conclusions from the published estimates.
Appendix 11. Implications of Population Controls on ACS Estimates

As the final step in the weighting process, ACS estimates are adjusted to conform with estimates from the Census Bureau’s Intercensal Population Estimates program. This adjustment is made at the sex, age, race, Hispanic origin, and total HUs level within county for the housing unit population, and at the state level total population by group quarters type for group quarters population. For single year estimates, the population and total housing unit estimates from July 1 of the same year are used as controls. For multiyear estimates, the controls are the average of the individual year intercensal estimates for each year in the multiyear period.

The primary issues associated with the use of the intercensal estimates is that these estimates require a number of national level assumptions and models defining changes in population and housing over time, in addition to direct information about the county. Thus, these estimates will tend to be insensitive to rapid growth or decline, and situations where local migration differs from migration for a larger geographic area. In addition, population estimates over the decade based on the previous census will not reflect fully the impact of seasonal changes in the basic demographics of the population of some areas across the calendar year.

The implication of these issues is that population controls may tend to dampen the effect of true population changes in small areas that are experiencing large population changes or which have population growth patterns which differ markedly from growth patterns in general. To the extent that population changes affect certain population subgroups (e.g., Hispanics), the accuracy of the ACS estimates could suffer.

Practitioners should also recognize that ACS single year estimates for 2010 and multiyear estimates with 2010 as the end year (e.g., 2008-2010) will use population controls data from the Intercensal Population Estimates program, which will most likely not agree with the 2010 Decennial Census population counts. The impact of this difference will likely be greater for smaller geographic areas where the potential for large differences between the population controls and the 2010 Decennial Census population counts is greater.
Appendix 12. Other ACS Resources

BACKGROUND & OVERVIEW INFORMATION

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

What is the ACS?
   This webpage 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
   This document describes the basic design of the 2005 ACS and details the full set of
   methods and procedures that were used in 2005.

ACS QUALITY

Accuracy of the Data (2006)
   This document provides data users with a basic understanding of the sample
design, estimation methodology, and accuracy of the 2006 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.

ACS Quality Measures
http://www.census.gov/acs/www/UseData/sse/
   This webpage includes information about the steps taken by the Census Bureau to
   improve the accuracy of ACS data. Four measures of survey quality are
   described and measures are provided at the state-level.
GUIDANCE ON DATA PRODUCTS AND USING THE DATA

2006 Data Users Handbooks: ACS
This handbook (in PDF format) provides detailed information about the ACS, including descriptions of the ACS content, data products, and quality measures. Examples are provided to illustrate how to access each data product.

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

Guidance on Comparing 2006 American Community Survey Data
http://www.census.gov/acs/www/UseData/compACS.htm
Tables are provided with guidance on comparing the 2006 ACS data products to 2005 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 Current Population Survey (CPS).