About ACS Estimates and Margins of Error in NEO CANDO 2010+

The data reported here are derived from the Census Bureau's American Community Survey (ACS) and require understanding and care in their use. The ACS is different from the decennial census and data users who are accustomed to using decennial census data should be aware of these differences.

One major difference is in the sampling error and associated margins of error associated with the ACS. This document provides a summary of the issues in using these data. In addition, because NEO CANDO 2010+ provides ACS-based data for locally unique geographic entities, such as neighborhoods, city wards, county council districts, and others, we describe the methods used to produce those data.

The Census Bureau publishes explanations of the ACS and its differences with the decennial census on its Web site [http://www.census.gov/acs/www/] and the reader is advised to consult this information when using the data. However, because of the importance of understanding the reliability of the ACS data we provide here a brief summary of the issues one must understand in using the data properly.[1]

About the American Community Survey and Differences with the Decennial Census
The ACS provides the ability to monitor social and economic trends in local communities on an annual basis. The ACS replaces the sample data that had been collected as part of the decennial censuses in previous decades.

Subjects
The subjects of the ACS are similar to those of the sample data in the 2000 census, including many socioeconomic and housing items census data with which many census data users are familiar (see Table 1 below, from page 1 of "What General Data Users Need to Know").

Table 1. **Subjects Included in the American Community Survey**

Demographic Characteristics
- Age
- Sex
- Hispanic Origin
- Race
- Relationship to Householder (e.g., spouse)

Economic Characteristics
- Income
- Food Stamps Benefit
- Labor Force Status
- Industry, Occupation, and Class of Worker
- Place of Work and Journey to Work
- Work Status Last Year
- Vehicles Available
- Health Insurance Coverage*

Social Characteristics
- Marital Status and Marital History*
- Fertility
- Grandparents as Caregivers
- Ancestry
- Place of Birth, Citizenship, and Year of Entry
- Language Spoken at Home
- Educational Attainment and School Enrollment
- Residence One Year Ago
- Veteran Status, Period of Military Service, and VA Service-Connected Disability Rating*
- Disability

Housing Characteristics
- Year Structure Built
- Units in Structure
- Year Moved Into Unit
- Rooms
- Bedrooms
- Kitchen Facilities
- Plumbing Facilities
- House Heating Fuel
- Telephone Service Available
- Farm Residence

Financial Characteristics
- Tenure (Owner/Renter)
- Housing Value
- Rent
- Selected Monthly Owner Costs

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

Methods of Collection
However, the method of collecting the data is very different, which introduces some difficulty in making comparisons and identifying trends over time. The three key differences between the 2000 Census sample data and ACS estimates are:

[1] This summary draws heavily on: 1) the Census Bureau's Web site; 2) the Census Bureau's "A Compass for Understanding and Using American Community Survey Data: What General Data Users Need to Know," U.S. Department of Commerce, Economics and Statistics Administration, U.S. Census Bureau, Washington, DC, October 2008; and 2) "The American Community Survey," an ESRI White Paper, April 2011.

1. Data collection - ACS data are collected in ongoing monthly surveys, while the 2000 Census was collected in a single, point-in-time survey (in April 2000).

2. Time frame – Because they are collected on a continuous basis (and because of the sample size issue noted below), the ACS data are reported for periods of time such as one year, three years, and five years periods, while the 2000 Census represents conditions as of approximately April 2000. Releases of ACS estimates include for all areas down to block groups and represent averages over 60 months, or five years.

3. Sample size –The ACS surveys one household in 40, while the 2000 Census was based on a 1 in 6 sample size. Both surveys introduce errors in estimating true values for the population (see discussion below) This issue was largely ignored with the use of the 2000 Census survey data but the smaller relative sample size for the ACS introduces significantly more concern for data users.

Some specific differences between the ACS and the 2000 Census estimates include:

- Residency rules are different. The ACS defines a resident by a two-month rule. The census rule is "usual place of residence" or wherever a person spends most of the year. ACS data may include seasonal populations in addition to year-round residents.

- Date-specific variables, like employment, represent monthly averages, including seasonal variations in the ACS.

- Since income, rent, and housing values and costs are collected over the course of the previous 12 months they must be adjusted by the Consumer Price Index to represent a calendar year.

- Migration is now measured from one year ago at the time of the survey, not five years ago as the 2000 Census asked respondents.

<u>1-Year, 3-Year, and 5-year Estimates</u>
The Census Bureau began the ACS program in 2000, developing its design and implementation over the first few years before full implementation in 2005. As shown in Table 2, starting with the data collected in 2005, ACS information has been published for areas with populations of 65,000 or more. In the fall of 2008, the first 3-year estimates were released based on aggregating data from the 2005, 2006, and 2007 surveys. In 2010, the ACS provided the first 5-year estimates of demographic, housing, social, and economic characteristics for the nation, states, cities, counties, census tracts, and block groups. These 5-year estimates are updated annually by removing the earliest year and replacing it with the latest one.

While a single-year estimate includes information collected over a 12-month period, a 3-year estimate represents data collected over a 36-month period, and a 5-year estimate includes data collected over a 60-month period. The Census Bureau suggests comparing periods that do not overlap—comparing 2005–2007 estimates with 2008–2010 estimates, for example. This means waiting longer to identify a trend. However, in areas undergoing fundamental shifts in the size or composition of the population, change may be so substantial that it will be obvious after only a few years.

The primary advantage of using multiyear estimates is the increased statistical reliability of the data for less populated areas and small population subgroups.

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

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

Year of Data Release: 2006–2013

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

Aside from definitional and data collection differences between the 2000 Census and the ACS, interpreting the change between the census taken April 2000 and a five-year average for 2006–2010 as provided in NEO CANDO 2010+ may be difficult because the ACS estimates cover a period of time that includes significant changes from beginning to end.[2]

Census Counts versus Survey Estimates

The primary purpose of the ACS is to measure the changing social and economic characteristics of the population. The ACS does not provide official counts of the population. Instead, the Census Bureau's Population Estimates Program is the official source for annual population totals, by age, race, Hispanic origin, and sex.[3] These population estimates are available for municipalities and counties, but not for census tracts or other small area geographies. ACS estimates are controlled to match the Census Bureau's annual population estimates, by age, sex, race, and Hispanic origin.

Because of the differences between counts and sample estimates, the Census Bureau advises ACS data users to use percentages and rates rather than estimates of the number of persons, households, or housing units produced by the ACS.

Geographic Areas Covered in the ACS

Table 3 shows the type and number of geographic areas included in the ACS, as well as the type of ACS estimates each will receive. The table is not exhaustive; it only covers major types of geographic areas. Geographic boundaries of some municipalities and townships change over time and are adjusted to most recent population estimates available at the time. Estimates from the 5-year, 2006-2010 ACS are based on census geography as of the 2010 Census. Only decennial census data are reported at the block level. ACS estimates are available for block groups and most other higher level geographies used by the Census Bureau.

NEO CANDO 2010+ includes estimates of the 2000 Census data for the 2010 census geography and for recent local geographies such as Cleveland's Statistical Planning Areas and others. These are not estimates provided by the Census Bureau, but rather by the Northeast Ohio Data Consortium (NEODC).[4]

[2] In 2006, the ACS began including samples of the population living in group quarters (e.g., jails, college dormitories, and nursing homes) for the first time. As a result, 2006 and later ACS data, including 3-year and 5-year estimates, may not be comparable with data from the 2005 ACS. This is especially true for estimates of young adults and the elderly, who are more likely than other groups to be living in group quarters facilities.

[3] For more information about population estimates, visit the Census Bureau's Web site at www.census.gov/popest/estimates.php.

[4] The NEODC includes The Center for Community Solutions, the Northern Ohio Data and Information Service (NODIS) in the Maxine Goodman Levin College of Urban Affairs at Cleveland State University, and the Center on Urban Poverty and Community Development in the Mandel School of Applied Social Sciences at Case Western Reserve University.

Table 3. Major Geographic Areas and Type of ACS Estimates Received

Type of geographic area	Total number of areas	Percent of total areas receiving . . .		
		1-year, 3-year, & 5-year estimates	3-year & 5-year estimates only	5-year estimates only
States and District of Columbia	51	100.0	0.0	0.0
Congressional districts	435	100.0	0.0	0.0
Public Use Microdata Areas*	2,071	99.9	0.1	0.0
Metropolitan statistical areas	363	99.4	0.6	0.0
Micropolitan statistical areas	576	24.3	71.2	4.5
Counties and county equivalents	3,141	25.0	32.8	42.2
Urban areas	3,607	10.4	12.9	76.7
School districts (elementary, secondary, and unified)	14,120	6.6	17.0	76.4
American Indian areas, Alaska Native areas, and Hawaiian homelands	607	2.5	3.5	94.1
Places (cities, towns, and census designated places)	25,081	2.0	6.2	91.8
Townships and villages (minor civil divisions)	21,171	0.9	3.8	95.3
ZIP Code tabulation areas	32,154	0.0	0.0	100.0
Census tracts	65,442	0.0	0.0	100.0
Census block groups	208,801	0.0	0.0	100.0

* When originally designed, each PUMA contained a population of about 100,000. Over time, some of these PUMAs have gained or lost population. However, due to the population displacement in the greater New Orleans areas caused by Hurricane Katrina in 2005, Louisiana PUMAs 1801, 1802, and 1805 no longer meet the 65,000-population threshold for 1-year estimates. With reference to Public Use Microdata Sample (PUMS) data, records for these PUMAs were combined to ensure ACS PUMS data for Louisiana remain complete and additive.

Source: U.S. Census Bureau, 2008. This tabulation is restricted to geographic areas in the United States. It was based on the population sizes of geographic areas from the July 1, 2007, Census Bureau Population Estimates and geographic boundaries as of January 1, 2007. Because of the potential for changes in population size and geographic boundaries, the actual number of areas receiving 1-year, 3-year, and 5-year estimates may differ from the numbers in this table.

[5] Public Use Microdata Areas, or PUMAs, are census-constructed geographic areas, each with approximately a population of 100,000. PUMAs do not cross state lines.

Locally Defined Geographies
For a description of the geographic entities used in NEO CANDO 2010 +, including both census geographies and locally defined geographies, click here.

Margin of Error
Survey-based estimates are subject to uncertainty about how well the estimates match reality. Two types of error can occur in survey data – sampling error and nonsampling error. Nonsampling errors can result from mistakes in how the data are reported or coded, or problems in the sampling frame or survey questionnaire.

Sample data from the 2000 Census represented a larger share of the population than does the ACS, and sampling errors were not reported. However, the Census Bureau deems it necessary to report measures of sampling error with all ACS estimates.

For any given area, the larger the sample (the more years included and/or larger the population from which the sample is drawn), the greater the confidence in the estimate. The estimate is more likely to accurately reflect the true value to be found in the entire population. The Census Bureau and NEO CANDO 2010+ report the margins of error for the 90-percent confidence level for ACS estimates. Ninety percent confidence intervals define a range expected to contain the true value of an estimate with a level of confidence of 90 percent. Margins of error define the upper and lower ends of these confidence ranges.

For example, the 2006-2010 ACS estimate of Cleveland's poverty rate, shown in Figure 1, is 31.2 percent. By adding and subtracting the margin of error from the point estimate, we calculate the 90-percent confidence interval for that estimate as follows:

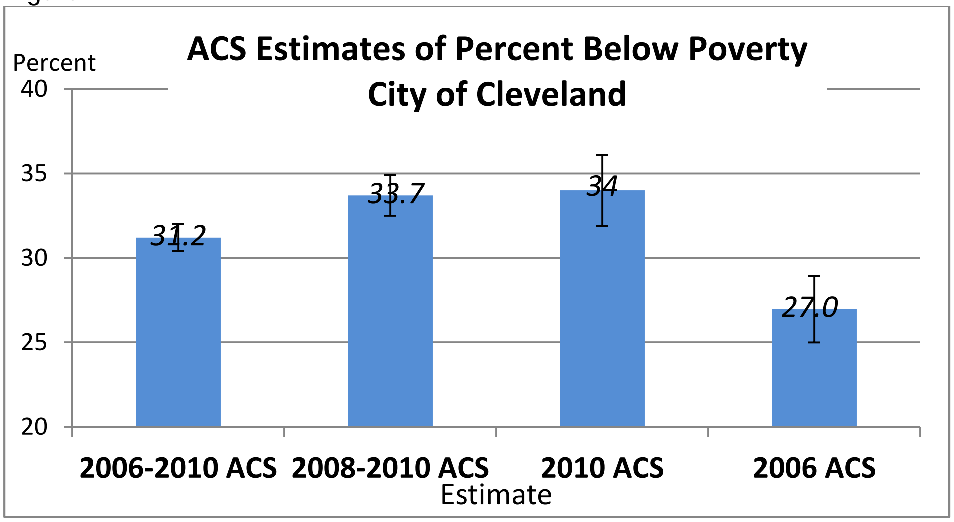
Lower-bound interval 31.2 – 0.8 = 30.4
Upper-bound interval 31.2 + 0.8 = 32.0

Therefore, we can be 90 percent confident that the true poverty rate for Cleveland during this period falls somewhere between 30.4 and 32.0. The I-bars in the graph illustrate this range.

Estimates and their margins of error for the 1-year 2010 and 2006 ACS estimates, the 3-year 2008-2010 ACS estimate, and the 5-year 2006-2010 estimate are also shown. Because the estimates are for different and overlapping time intervals it is difficult to say precisely how the poverty rate changed from 2006 to 2010. Yet the estimates clearly suggest that it increased over this period.

Indeed, using the 1-year estimates for 2006 and 2010 we can conclude that the poverty rate probably increased from 2006 to 2010, since the lower bound in the 2010 estimate is 31.9 percent, while the upper bound in the 2006 estimate is 28.9 percent (based on a margin of error of 2.0). We are 90 percent confident that the true poverty rate in 2010 was higher than the rate in 2006.[5]

Figure 1



How Estimates and MOEs for Local Geographies Were Calculated
Estimating the ACS data for geographic locally developed summery levels not provided by the Census Bureau, such as SPAs, Cuyahoga County council districts, Cleveland wards, and Summit County

[5] Overlapping confidence intervals and statistical tests of significance are related but not the same thing. A relatively simple description of this issue is provided in Knezevic, Andrea, "Overlapping Confidence Intervals and Statistical Significance," StatNews Number73, Cornell Statistical Consulting Unit, Cornell University, October 2008 [http://www.cscu.cornell.edu/news/statnews/stnews73.pdf] in which it is concluded that "Generally, when comparing two parameter estimates, it is always true that if the confidence intervals do not overlap, then the statistics will be statistically significantly different. However, the converse is not true." That is, overlapping intervals may still occur when the two estimates are statistically different. A more technical discussion is provided by Payton, Mark E., Mathew H. Greenstone, and Nathaniel Schenker, "Overlapping confidence intervals or standard error intervals: What do they mean in terms of statistical significance?" in Journal of Insect Science, 3:34, October 2003 [http://www.ncbi.nlm.nih.gov/pmc/articles/PMC524673/].

Clusters, required calculation of MOEs for these data. We used the Bureau's proscribed methods. See the section on Calculating MOEs below.

Locally defined area estimates and MOEs are built using ACS base tables available for county subdivisions, places, census tracts and block groups, depending on how the local geographies are composed. We used the largest (highest) summary level where possible so that estimated MOEs are based on the largest sample size available. In addition, using fewer geographic entities in aggregating estimates produces smaller margins of error. Nevertheless, local geography boundaries are often split by the Census geographies, requiring use of block group level estimates and MOEs.

We use the base tables rather than extracting estimates from the Census Bureau's American FactFinder because in many cases we needed block group summary level data to estimate the data for locally defined geographies. Block group level estimates are not available in the American FactFinder. Furthermore, the base tables include more tables of estimates than those provided in American FactFinder. However, not all ACS data are available at the block group level and in those cases tract level estimates are used.

The method of calculating MOEs when aggregating count estimates calls for summing the squared values of the individual MOEs for the units involved (such as block groups when aggregating to a larger geographic area). However, because some locally defined geographic areas share one or more block groups, we had to first estimate the portions of the estimated block group count that belong to the areas sharing them. In such cases we used the 2010 population or housing units of the census blocks that compose the block group to apportion the ACS estimates and squared MOEs to the parts of the block group shared by the locally defined areas. In the cases where a ACS 2006-2010 block group has a reported population or housing units, but the 2010 block group does not, the geographic area of the block group in each of the locally defined areas sharing it is used to proportion the data.

After making such adjustments the data (both count estimates and their associated squared MOEs) are aggregated. However, as proscribed by the Census Bureau, when there is more than one estimate of zero among the count estimates being aggregated, only one of the MOEs is used.[6]

Percentage and ratio estimates are recalculated based on the counts estimates. See methods described in the section on Calculating MOEs below. [7]

Estimated medians are recalculated by interpolating within the range containing the 50th percentile. There is no optimal method to calculate margin of errors for a median, so a solution obtained from the Census Bureau using a variation of the PUMS design factor-based methodology is used.[8]

---

[6] The issue with MOEs and zero estimates is described in the American Community Survey Multiyear Accuracy of the Data (3-year 2008-2010 and 5-year 2006-2010) http://www.census.gov/acs/www/Downloads/data_documentation/Accuracy/MultiyearACSAccuracyofData2010.pdf on page 21 A.

[7] The method for calculating MOEs for derived proportions (or rates) is described on A14-A15 of A Compass for Understanding and Using American Community Survey Data: http://www.census.gov/acs/www/Downloads/handbooks/ACSGeneralHandbook.pdf.

[8] The design factor-based methodology is described on pages 16-17 of the ACS PUMS Accuracy of the Data document at this URL: http://www.census.gov/acs/www/Downloads/data_documentation/pums/Accuracy/2010AccuracyPUMS.pdf. Instead of using 99 (a finite population correction factor which is appropriate for the 1-year PUMS files), for the 5-year full sample ACS data 7 was used.

Please note that calculated margins of error vary depending on the number of fields used and the number of geographies used. Also, areas with very small population will tend to have large margin of errors because of the ACS sample size.[9]

Calculating MOEs

The Census Bureau provides MOEs for its estimates and so does NEO CANDO 2010+. However, there are times when the user must calculate MOEs. This includes situations in which the user aggregates estimates for geographic areas, aggregates estimates for new categories of data (such as groupings of ages), or when calculating new variables such proportions, percentages, and rates that are not otherwise provided.

While Census and NEO CANDO 2010+ estimates should cover most types of geographic units needed by data users in northeast Ohio, there may be times when the user will want to aggregate estimates for custom geographic areas. Aggregating census tract-level estimates for a user-defined service area, for example, will require the user to calculate MOEs in order to evaluate the reliability of the estimate.

Aggregating ACS estimates within a table of estimates, such as custom groupings of occupations or educational attainment, or for ranges of years of housing construction, income, and others. Percentage or rate calculations, if not already provided by the Census Bureau or NEO CANDO 2010+, will also require calculation of MOEs to assess the reliability of the estimates.

<u>For aggregating count data into new categories or geographic areas:</u>

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

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

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

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

[9]Some of these issues are described in the ACS Accuracy of the Data in section Issues with approximating the standard error of linear combinations of multiple estimates B through E, starting on page 22. The URL is http://www.census.gov/acs/www/Downloads/data_documentation/Accuracy/MultiyearACSAccuracyofData2010.pdf

For derived proportions or percentages

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

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

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

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

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

MOE_{den} is the MOE of the denominator.

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

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

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

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

<u>For derived ratios</u>

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

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

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

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

where MOE_{mum} 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.

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

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

Avoiding Pitfalls When Working With ACS Data

Below are some key mistakes to avoid and some precautions:

- Use caution in comparing ACS data with data from the decennial census or other sources. Every survey uses different methods, which could affect the comparability of the numbers.
- The ACS was designed to provide estimates of the characteristics of the population, not to provide counts of the population in different geographic areas or population subgroups.
- Be careful in drawing conclusions about small differences between two estimates because they may not be statistically different.
- Data users need to be careful not to interpret annual fluctuations in the data as long-term trends.

- Data users should not interpret or refer to 3-year or 5-year period estimates as estimates of the middle year or last year in the series. For example, a 2005–2007 estimate is not a “2006 average.”

- While it is safe to say that non-overlapping confidence intervals mean that the estimates are statistically different (i.e., are estimates from different populations or groups), overlapping intervals do not necessarily mean that the estimates represent the same population.

Finally, some users may want to use a confidence interval other than the 90 percent level being used by the Census Bureau and NEO CANDO. This can be accomplished by multiplying the published MOE by an adjustment factor. See pages A-12 and A-13 of “A Compass for Understanding and Using American Community Survey Data” for details.[10]

[10] http://www.census.gov/acs/www/Downloads/handbooks/ACSGeneralHandbook.pdf