



**UNITED STATES DEPARTMENT OF COMMERCE**  
**Economics and Statistics Administration**  
**U.S. Census Bureau**  
Washington, DC 20233-0001

DSSD 2008 AMERICAN COMMUNITY SURVEY RESEARCH MEMORANDUM SERIES ACS08-R-05

MEMORANDUM FOR ACS Research and Evaluation Team

From: David C. Whitford *signed 5/13/11*  
Chief, Decennial Statistical Studies Division

Prepared by: Michael Starsinic and Karen E. King  
ACS Variance Estimation and Statistical Support Branch  
Decennial Statistical Studies Division

Subject: Quality Rating Classification of 5-year ACS MYES Estimates by Population  
Size Groupings and in Comparison with Census 2000 Long Form Estimates

The document attached summarizes results of two American Community Survey (ACS) Research and Evaluation projects on the potential quality or reliability of ACS 5-year data undertaken at the request of Director Steve H. Murdock in the fall of 2008. These projects were "Quality Rating Classification of 5-year ACS Estimates by Population Size Groupings" and "Comparing the Quality of 5-year ACS Estimates to Census 2000 Long Form Estimates". The only 5-year ACS estimates available at the time of this research were from the Multiyear Estimate Study (MYES) conducted in 2006, which included data for 34 of the ACS test counties and covered the period from 1999 through 2005. Census 2000 Long Form data came from the ACS/Census Comparison Project. These results were presented to the director on November 20, 2008.

If you have questions regarding it, please contact Michael Starsinic at (301)763-3606 or Karen King at (301)763-1974.

Attachment

Cc:  
A. Navarro (DSSD)  
A. Tersine  
J. Hartman  
J. Karberg  
S. Hefter  
M. Asiala  
D. Griffin (ACSO)  
L. Weidman (SRD)

# Quality Rating Classification of 5-year ACS MYES Estimates by Population Size Groupings and in Comparison with Census 2000 Long Form Estimates

FINAL REPORT



MICHAEL STARSINIC  
ALFREDO NAVARRO  
KAREN KING  
DECENNIAL STATISTICAL STUDIES DIVISION

## Quality Rating Classification of Estimates from the 2001-2005 ACS Multiyear Estimate Study by Population Size and in Comparison with Census 2000 Long Form Estimates

by Michael Starsinic, Alfredo Navarro, and Karen King  
Decennial Statistical Studies Division

### Introduction

In 2010, the American Community Survey (ACS) began releasing all 5-year estimates for census tracts, census block groups, and small governmental units without any data quality filtering. However, concerns were raised prior to that about the reliability of many of the estimates likely to be included in the data products. Based on a request by Director Steve H. Murdock in the fall of 2008, a series of assessments were conducted to look at the possible quality or reliability of 5-year ACS data. The only 5-year ACS estimates available in 2008 were from the Multiyear Estimate Study (MYES) conducted in 2006, which included data for 34 of the ACS test counties and covered the period from 1999 through 2005. The assumption is that the patterns of quality seen in these data will be consistent with the quality of the first 5-year estimates released in 2010. This report summarizes results of two of the assessments undertaken that were presented to the director on November 20, 2008.

The ACS has been described as the replacement for the Census Long Form. The ACS will ultimately release multiyear estimates for the same small areas that the Long Form did. The quality of the Census Long Form estimates is therefore an important benchmark for comparing projected 5-year reliability measures for the ACS to comparable Census 2000 Long Form measures.

The error in survey estimates typically includes two components: sampling and nonsampling error. This assessment is limited to aspects of sampling error as measured by the coefficient of variation (CV). The CV is a relative measure of reliability defined as the ratio of the standard error of an estimate to the estimate itself. For example, a 10 percent CV on an estimate of a 20 percent population characteristic suggests a standard error equal to 2 percent.

This assessment will address three questions surrounding the issue of the reliability of ACS estimates for small areas.

- Question 1 – Is the reliability of ACS multiyear estimates adequate?
- Question 2 – Does the reliability of ACS estimates vary by size of area and topic?
- Question 3 – How does the ACS reliability compare to the Census 2000 Long Form?

To answer the first two questions, we quantified the reliability of ACS 5-year estimates for small areas according to specific rules based on the CV, resulting in a classification of “High”, “Medium”, or “Low”. Results were summarized for a subset of estimates from the Data Profiles by size of area.

Similarly, results were summarized by group of estimates usually referred to as subject or topic areas. Results are displayed for subjects such as race (major groups) and housing tenure.

To answer the third question we quantified the reliability of the 2001-2005 multiyear estimates from the MYES and compared it to the Census 2000 Long Form estimates for the same areas. The comparison is based on the average CV for the estimate's distribution. Results are also displayed for selected profile estimates.

## Methodology

All ACS data used in this research were produced as part of the ACS MYES. The MYES used data collected in 34 ACS test counties for the years 1999 through 2005 and produced a series of 1-year, 3-year, and 5-year period estimates. This study allowed the Census Bureau to test possible techniques for the production of multiyear ACS estimates and release a reduced set of data products on the ACS website for public use. For the remainder of this document estimates from the ACS MYES will be referred to as ACS estimates.

Specifically, this research uses detailed tables and data profiles for the 2001-2005 5-year period estimates. These estimates include the "g-weighting" methodology, first employed in the MYES, to reduce standard error estimates for small areas. These estimates do not include any group quarters (GQ) data, as the ACS collected no GQ data until 2006.

For the 2001-2005 ACS MYES 5-year products, detailed tables and data profiles were produced for a total of 10,906 geographic areas within these 34 counties, including places, minor civil divisions (MCDs), American Indian and Alaska Native (AIAN) areas, zip code tabulation areas (ZCTAs), public use microdata areas (PUMAs), school districts (SDs), tracts, and block groups. Data for the 6,913 block groups are excluded from this analysis. Table 1 shows the breakdown of the 3,993 geographic areas used in this study by summary level and population size.

Table 1: Number of 2001-2005 MYES Geographic Areas, By Summary Level and Pop Size

Type of Area		Number of Areas							
		Total	< 20,000	<1,000	1,000 to 4,999	5,000 to 9,999	10,000 to 19,999	20,000 to 65,000	65000+
050	county	34	4	0	0	1	3	11	19
060	MCD	373	302	87	146	34	35	48	23
140	tract	2,270	2,268	75	1,429	725	39	2	0
160	place	476	415	149	164	64	38	42	19
250	AIAN area	4	4	1	3	0	0	0	0
795	PUMA	71	0	0	0	0	0	0	71
871	ZCTA	531	348	118	97	53	80	166	17
950	Elementary SD	96	87	20	38	15	14	9	0
960	Secondary SD	19	7	0	0	2	5	10	2
970	Unified SD	119	67	4	16	19	28	35	17
	Total by size	3,993	3,502	454	1,893	913	242	323	168

About 88 percent of all areas used in this research are under 20,000 in population size, meaning they would only be published in 5-year ACS data products. Another 8 percent are between 20,000 and 65,000, and the remaining 4 percent are above 65,000 in population.

The Census 2000 Long Form estimates used in this research are also taken from a previous ACS research project. In the **1999-2001 ACS and Census 2000 Long Form Comparison Study**, three years of ACS sample data were combined for the 34 ACS test counties and compared at the county and tract level against Census 2000 Long Form estimates. The long form estimates for this project were specially tabulated excluding GQ persons, to make the data more comparable with the ACS data, which also was without GQ persons. ACS-like data profiles were prepared at the tract level for the long form data without GQ persons, and standard errors (using the design factor methodology available to the public) were calculated for each estimate. This long form profile data was re-used for this research project.

For this research, we wanted to create a quality indicator flag that could be applied not to an estimate for a specific geographic area, but instead to a line (or a group of lines) in a table or profile across geographic areas. It wouldn't tell a user "what is the quality of this estimate for this specific geographic area", but it would answer "what is the usual quality of this estimate?" We classified an estimate in one of three "quality rating" categories:

1. "High", if more than 80 percent of all the estimates for all geographic areas have CVs less than 0.3.
2. "Med", for the medium range, if between 60 and 80 percent of all geographic areas have CVs for the estimate less than 0.3.
3. "Low", if less than 60 percent of all the estimates for all geographic areas have CVs less than 0.3.

## Results

### Question 1– Is the reliability of ACS multiyear estimates adequate?

Table 2: Percent of MYES Profile Estimates By Quality Rating Category and Population Size Groups Up to 20,000.

Population Size	< 1,000		1,000 to 4,999		5,000 to 9,999		10,000 to 20,000	
Number of Areas	454		1,893		913		242	
Total Number of Estimates	433	100%	433	100%	433	100%	433	100%
"High" Quality Estimates	17	3.9%	107	24.7%	145	33.5%	212	49.0%
"Med" Quality Estimates	44	10.2%	30	6.9%	48	11.1%	39	9.0%
"Low" Quality Estimates	372	85.9%	296	68.4%	240	55.4%	182	42.0%

Table 2 shows the Quality Rating over all count estimates (of persons, households, and housing units) in the MYES data profile by size of area, for the areas up to 20,000. We see for areas with population between 5,000 and 9,999, there are 145 of the 433 with a rating of "High". Reader should note that there are actually 454 count estimates in each MYES data profile, but 21 are duplicates of other lines and are not included in Table 2.

As an example, let us look at a specific estimate such as the estimate of population 21 years and older. We calculated the CV for the estimate of population 21+ for each of the 913 geographic areas in the 5,000-9,999

population size category, and determined the percentage of areas with CV less than 0.3. Looking at Table 3, given page 7, we see that 100% of the areas have a CV less than 0.3, so, by our criteria, this estimate is considered to be of “High” reliability for this size group. Therefore, the population 21+ and 144 other estimates are of “High” quality for areas with total population between 5,000 and 9,999.

In Table 3, we see the overall Quality Rating for selected MYES profile estimates by size of area and percent of these sized areas with CV under 0.3 for that item. The quality ratings in the first four columns can be derived from the percent of areas with a CV less than 0.3 in the last four columns. Estimates in a size category with less than 60% will be “Low”, between 60% and 80% will be “Med” and above 80% will be “High”.

The 16 estimates displayed in Table 3 were selected to be essentially representative of the many estimates in the profile, including estimates of high interest and crossing a wide range of topics. To not give a false impression of quality, several small or rare characteristics, such as mobile homes, were included as well.

Continuing working with our estimate of the 21+ population for areas in the 5,000 to 9,999 population range, we see that “High” is the quality rating. Looking to the right, we see why this estimate is given this rating. By our criteria “High” is given when at least 80% of the areas had a CV less than 0.3. In this case, we see that 100.0% of the areas had CVs less than 0.3. For larger areas the percent of areas with CV smaller than 0.3 is 100.0% as well. Whereas we see for the population older than 65 has a rating of “Low” because only 31.1% of areas with less than 1,000 population have a CV less than 0.3.

#### Question 2– Does the reliability of ACS estimates vary by size of area and topic?

Table 4, given on page 8, shows a quality rating for the overall profile topic level by size of area up to 20,000. Here, we are looking across all lines in the data profile that fit into the topic, not at individual lines. The percent of areas with the CV less than 0.3 is calculated in the same way, but by looking across both dimensions: number of geographic areas and line estimates within a topic.

Working with all the age/sex estimates, of which 21+ population and 65+ population are two of the 31 age/sex estimates in the profile, for areas in the 5,000 to 10,000 population size category, we see “High” as the quality rating. If we look to the right, we see why this profile topic is given this rating. By our criteria “High” is given when more than 80% of all age and sex estimates within all areas in a given population size category have a CV less than 0.3. In this case, we see 92.0% of the estimates in this topic within these areas had CVs less than 0.3. As we go into larger areas we see the percent increases to 97.9%.

We can also see the impact of using the topic groupings instead of individual lines when applying the rating. For example, utility gas in Table 3 was rated “Med” and “High” for areas of 5,000-9,999 and 10,000-20,000 population, respectively. However, the heating fuel topic in Table 4 is “Low” for both those size groups and the percent of areas with CVs less than 0.3 are not close to 80%. This is because the topic includes several lines with small or rare characteristics, including “solar” and “coal or coke”, which have poor reliability almost everywhere. The point here is that this is a function of the size of the estimate and not the result of sample design inefficiencies.

### Question 3 - How does the ACS reliability compare to the Census 2000 Long Form?

For Tables 5 and 6, given on pages 9 and 10, we are shifting from all areas up to 20,000 to focus only on census tracts. The long form data used here, as described in the methodology section, was specially created for the ACS/Census comparison project, and consists of household data only. Tabulations of this special long form product were only produced for the ACS test counties and tracts within those counties. Using the 1999-2001 ACS and comparison study data, we can directly compare ACS and Census 2000 Long Form estimates and CVs. For simplicity, we will continue examining the same set of 16 estimates as was used in Table 3. Table 5 assigns the quality rating classification to both ACS and Census 2000 Long Form estimates.

Continuing working with our estimate of population 21+ for census tracts with less than 1,000 population and using the same criteria for quality rating, we see the “High” as the quality rating for both ACS and Census 2000 Long Form. This is not surprising since we saw this earlier in Table 3 for ACS and we expect the Census 2000 Long Form to have smaller CVs than the ACS.

We see again the 65+ population has a rating of “Low” in census tracts with a population below 1,000. The Census 2000 Long Form rating for the 65+ population is low as well. Note in Table 3, the 65+ population estimate was given a rating of “Med” for areas 1,000 – 4,999, but Table 5 shows a rating of “High” for this size area. The difference is that Table 3 shows results for all level of geography where as Table 5 shows results for census tracts only.

Table 6 shows the average CV for selected estimates by size of tract in occupied housing units for ACS and Census 2000 Long Form.

As expected, the Long Form mean CVs are smaller than the ACS mean CVs. Across all tracts, the ACS mean CVs are frequently about 60% higher than the long form CVs, which is broadly consistent with previous research.

### **Main Observations**

- The quality rating system introduced here can be used to classify table and profile lines as having high, medium or low quality.
- For small areas of geography up to 20,000, a significant number of ACS MYES estimates have a “Med” or “High” quality rating based on the criteria used that at least 60% of areas had an estimated CV less than 0.3. For areas between 10,000 and 20,000 population the majority of these estimates get a quality rating of “Med” or “High”.
- Some estimates are of “High” or “Med” quality in all the size categories such as 21+ population and percent of high school graduates or higher.
- Several overall profile topics have a rating of “Low” in all size categories up to 20,000 population. Again, this is a function of including dictated small size characteristic estimates in the data profiles.
- Using the quality ratings, many 1999- 2001 ACS and Census 2000 Long Form estimates have the same rating in the same size categories. Some long form estimates do get assigned a “Low” rating which indicates not all long form estimates are great quality.
- Differences observed in mean CVs between these ACS and Census 2000 Long Form estimates are overall as expected.

Table 3: Quality Rating for Selected Profile Estimates, by Population Size Group (Areas of population up to 20,000)

Profile Estimate	Quality Rating				Percent of areas with CV < 0.3			
	Population Size (in thousands)				Population Size (in thousands)			
	< 1	1 - < 5	5 - <10	10 - 20	< 1	1 - <5	5 - <10	10 - 20
<b>SEX AND AGE</b>								
21 years and over	Med	High	High	High	72.9%	100.0%	100.0%	100.0%
65 years and over	Low	High	High	High	31.1%	89.4%	98.5%	100.0%
<b>RACE</b>								
Asian	Low	Low	Low	Low	0.4%	9.0%	19.4%	33.9%
<b>HOUSEHOLDS BY TYPE</b>								
Married-couple families	Low	High	High	High	52.4%	94.9%	99.8%	100.0%
<b>EDUCATIONAL ATTAINMENT</b>								
Less than 9th grade	Low	Low	Low	Med	4.2%	28.5%	50.1%	66.1%
Percent high school graduate or higher	High	High	High	High	86.1%	99.8%	100.0%	100.0%
Percent bachelor's degree or higher	Low	Med	High	High	21.4%	78.9%	92.4%	99.2%
<b>PLACE OF BIRTH</b>								
State of residence	Low	High	High	High	58.6%	98.7%	99.9%	100.0%
<b>EMPLOYMENT STATUS</b>								
Not in labor force	Low	High	High	High	55.1%	99.6%	100.0%	100.0%
<b>COMMUTING TO WORK</b>								
Car, truck, or van -- carpoled	Low	Low	Med	High	1.8%	25.7%	60.6%	93.4%
<b>INDUSTRY</b>								
Retail trade	Low	Low	High	High	1.5%	44.8%	82.0%	98.3%
<b>INCOME AND BENEFITS (IN 2005 INFLATION-ADJUSTED DOLLARS)</b>								
With Social Security	Low	High	High	High	39.4%	92.6%	98.7%	100.0%
<b>PERCENTAGE OF FAMILIES AND PEOPLE WHOSE INCOME IN THE PAST 12 MONTHS IS BELOW THE POVERTY LEVEL</b>								
All people	Low	Low	Med	High	9.9%	50.5%	65.4%	84.7%
<b>UNITS IN STRUCTURE</b>								
Mobile home	Low	Low	Low	Low	8.4%	20.2%	26.0%	45.0%
<b>HOUSING TENURE</b>								
Renter-occupied	Low	High	High	High	17.8%	82.6%	92.9%	98.8%
<b>HOUSE HEATING FUEL</b>								
Utility gas	Low	Med	Med	High	22.0%	78.2%	76.3%	83.9%

Table 4: Quality Rating for Selected Profile Topic, by Population Size Group (Areas of population up to 20,000)

Profile Topic	Quality Rating				% of areas with CV < 0.3			
	Population Size (in thousands)				Population Size (in thousands)			
	< 1	1 - <5	5 - <10	10- 20	< 1	1 - < 5	5 - <10	10- 20
<b>SEX AND AGE</b>	Low	High	High	High	38.1%	80.7%	92.0%	97.9%
<b>RACE (major group)</b>	Low	Low	Low	Low	11.3%	23.9%	30.3%	35.0%
<b>HOUSEHOLDS BY TYPE</b>	Low	Med	Med	High	21.9%	65.3%	76.8%	88.1%
<b>EDUCATIONAL ATTAINMENT</b>	Low	Med	High	High	23.1%	66.1%	82.9%	93.9%
<b>PLACE OF BIRTH</b>	Low	Med	High	High	36.6%	72.7%	81.3%	85.7%
<b>EMPLOYMENT STATUS</b>	Low	Med	High	High	37.4%	76.1%	85.3%	92.5%
<b>COMMUTING TO WORK</b>	Low	Low	Low	Low	19.6%	36.3%	43.5%	58.3%
<b>INDUSTRY</b>	Low	Low	Low	Med	2.1%	23.4%	44.5%	78.5%
<b>INCOME AND BENEFITS (IN 2005 INFLATION-ADJUSTED DOLLARS)</b>	Low	Low	Med	High	32.0%	57.4%	68.6%	81.7%
<b>PERCENTAGE OF PEOPLE WHOSE INCOME IN THE PAST 12 MONTHS IS BELOW THE POVERTY LEVEL</b>	Low	Low	Low	Low	15.1%	29.6%	41.1%	58.6%
<b>UNITS IN STRUCTURE</b>	Low	Low	Low	Low	10.2%	24.7%	32.7%	50.0%
<b>HOUSING TENURE</b>	Low	High	High	High	40.3%	87.9%	93.1%	99.0%
<b>HOUSE HEATING FUEL</b>	Low	Low	Low	Low	8.9%	21.7%	23.1%	31.4%

Table 5: Quality Rating for Selected Profile Estimates at the Tract Level, 1999-2001 ACS vs. Census 2000 Long Form, by Population Size Group

Profile Estimate	Quality Rating - ACS				Quality Rating - Long Form			
	Population Size (in thousands)				Population Size (in thousands)			
	< 1	1 - <5	5 - 10	>10	< 1	1 - <5	5 - 10	>10
<b>SEX AND AGE</b>								
21 years and over	High	High	High	High	High	High	High	High
65 years and over	Low	High	High	High	Low	High	High	High
<b>RACE</b>								
Asian	Low	Low	Low	Low	Low	Low	Low	Med
<b>HOUSEHOLDS BY TYPE</b>								
Married-couple families	Low	High	High	High	Med	High	High	High
<b>EDUCATIONAL ATTAINMENT</b>								
Less than 9th grade	Low	Low	Low	Low	Low	Med	High	High
Percent high school graduate or higher	High	High	High	High	High	High	High	High
Percent bachelor's degree or higher	Low	High	High	High	Low	High	High	High
<b>PLACE OF BIRTH</b>								
State of residence	Low	High	High	High	High	High	High	High
<b>EMPLOYMENT STATUS</b>								
Not in labor force	Med	High	High	High	High	High	High	High
<b>COMMUTING TO WORK</b>								
Car, truck, or van -- carpoled	Low	Low	Low	High	Low	High	High	High
<b>INDUSTRY</b>								
Retail trade	Low	Low	High	High	Low	High	High	High
<b>INCOME AND BENEFITS (IN 2005 INFLATION-ADJUSTED DOLLARS)</b>								
With Social Security	Low	High	High	High	Low	High	High	High
<b>PERCENTAGE OF FAMILIES AND PEOPLE WHOSE INCOME IN THE PAST 12 MONTHS IS BELOW THE POVERTY LEVEL</b>								
All people	Low	Low	Med	Med	Low	High	High	High
<b>UNITS IN STRUCTURE</b>								
Mobile home	Low	Low	Low	Low	Low	Low	Low	Low
<b>HOUSING TENURE</b>								
Renter-occupied	Low	High	High	High	High	High	High	High
<b>HOUSE HEATING FUEL</b>								
Utility gas	Low	High	Med	Low	Med	High	High	Med

Table 6: Average CV of Selected 1999-2001 ACS and Census 2000 Long Form Estimates, By Number of Occupied Housing Units in Tract

	Source	Tracts w/ <2000 OHU	Tracts w/ 2000- 4000 OHU	Tracts w/ >4000 OHU	All Tracts
<b>SEX AND AGE</b>					
21 years and over	ACS	0.042	0.029	0.020	0.038
	LF	0.033	0.021	0.014	0.029
65 years and over	ACS	0.198	0.138	0.101	0.179
	LF	0.161	0.113	0.075	0.147
<b>RACE</b>					
Asian	ACS	0.643	0.505	0.360	0.593
	LF	0.936	0.805	0.669	0.892
<b>HOUSEHOLDS BY TYPE</b>					
Married-couple families	ACS	0.145	0.103	0.074	0.132
	LF	0.086	0.061	0.051	0.079
<b>EDUCATIONAL ATTAINMENT</b>					
Less than 9th grade	ACS	0.444	0.374	0.386	0.422
	LF	0.342	0.285	0.317	0.326
High school graduate or higher	ACS	0.053	0.034	0.018	0.047
	LF	0.034	0.020	0.013	0.030
Bachelor's degree or higher	ACS	0.220	0.140	0.081	0.194
	LF	0.147	0.083	0.054	0.128
<b>PLACE OF BIRTH</b>					
State of residence	ACS	0.086	0.082	0.103	0.085
	LF	0.057	0.054	0.102	0.057
<b>EMPLOYMENT STATUS</b>					
Not in labor force	ACS	0.116	0.089	0.065	0.107
	LF	0.073	0.054	0.041	0.067
<b>COMMUTING TO WORK</b>					
Car, truck, or van -- carpooled	ACS	0.400	0.320	0.271	0.374
	LF	0.238	0.172	0.156	0.218
<b>INDUSTRY</b>					
Retail trade	ACS	0.339	0.244	0.186	0.309
	LF	0.227	0.149	0.130	0.204
<b>INCOME AND BENEFITS (IN 2005 INFLATION-ADJUSTED DOLLARS)</b>					
With Social Security	ACS	0.184	0.132	0.097	0.167
	LF	0.177	0.122	0.086	0.161

OHU – Occupied Housing Units. LF – Census 2000 Long Form Estimates.

		Tracts w/	Tracts w/	Tracts w/	
	Source	< 2000 OHU	2000- 4000 OHU	>4000 OHU	All Tracts
<b>PERCENTAGE OF FAMILIES AND PEOPLE WHOSE INCOME IN THE PAST 12 MONTHS IS BELOW THE POVERTY LEVEL</b>					
All people	ACS	0.339	0.268	0.237	0.317
	LF	0.214	0.144	0.121	0.194
<b>UNITS IN STRUCTURE</b>					
Mobile home	ACS	0.547	0.533	0.705	0.546
	LF	0.807	0.929	1.352	0.851
<b>HOUSING TENURE</b>					
Renter-occupied	ACS	0.166	0.112	0.100	0.149
	LF	0.116	0.069	0.060	0.102
<b>HOUSE HEATING FUEL</b>					
Utility gas	ACS	0.154	0.220	0.279	0.175
	LF	0.137	0.254	0.274	0.170

OHU – Occupied Housing Units. LF – Census 2000 Long Form Estimates.