



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

DSSD 2010 AMERICAN COMMUNITY SURVEY MEMORANDUM SERIES #ACS10-R-4

SEP 29 2010

MEMORANDUM FOR David C. Whitford
Chief, Decennial Statistical Studies Division

From: Alfredo Navarro *AN*
Assistant Division Chief, ACS Statistical Studies

Prepared by: Daniel Sommers *DS* and Steven P. Hefter *SPH*
American Community Survey Sample Design Branch
Decennial Statistical Studies Division

Subject: American Community Survey Sample Stratification - Current and
New Methodology

The attached handout, American Community Survey (ACS) Sample Stratification - Current and New Methodology, was distributed at the meeting of the ACS Research Steering Committee on June 24, 2010.

Any questions regarding this handout should be directed to Daniel Sommers at (301) 763-3714 or Steven Hefter at (301) 763-4082.

Attachment

cc:

DSSD Management Group
ACSO Managers Meeting List
ACSSDB

American Community Survey Sample Stratification - Current and New Methodology

The ACS produces estimates for geographic areas having a wide range of population sizes. To ensure that the estimates for these areas have the desired level of reliability, areas with smaller populations must be sampled at higher rates relative to those areas with larger populations.

Currently, the ACS stratification methodology assigns each block and its constituent addresses to one of seven sampling strata, each with a unique sampling rate. Six of the seven sampling rates are a function of the base rate which is recalculated each year. The rate for the smallest stratum has been fixed at 10 percent since 2004. The assignment of a block to a sampling stratum is based on information about the set of governmental units which contain the block, or on information about the size of the census tract that the block is located in¹. In order to determine the base rate, the distribution of addresses by sampling stratum and the target sample size, along with each stratum's rate definition, are used to solve a simple algebraic equation.

The five current sampling strata and their sampling rates are shown in Table 1.

Table 1. Current Stratification and Annual Sampling Rates

Stratum	Block MOS ² Criteria	Annual Sampling Rate
1	800 <= GUMOS <= 1200	1.5 x Base Rate
2	200 <= GUMOS < 800	3 x Base Rate
3	0 <= GUMOS < 200	0.10
4	0 <= TRACTMOS < 2000	Base Rate
5	0 <= TRACTMOS < 2000 High Response	0.92 x Base Rate
6	2000 <= TRACTMOS	0.75 x Base Rate
7	2000 <= TRACTMOS High Response	0.92 x 0.75 x Base Rate

In 2007, the National Academies of Science Committee on National Statistics (NAS, 2007) made the following recommendation to the Census Bureau:

“Recommendation 4-4: The Census Bureau should identify potential ways to increase the precision of ACS estimates for small geographic areas, particularly small governmental jurisdictions, through reallocation of the sample and through increases in the overall sample size. Cost savings should be sought to support such increases, although increases that could significantly improve the precision of estimates will require additional funding from Congress. Sample reallocation should also be considered to minimize anomalies across areas (for example, jurisdictions with very similar populations that fall into different sampling rate categories)”

¹ Governmental units include Counties, Places (active functioning governmental units), School Districts, American Indian Areas, Tribal Subdivisions (active, functioning), Minor Civil Divisions (in the twelve strong MCD states), Alaskan Native Village Statistical Areas, Hawaiian Homelands, and Census designated Places (Hawaii only).

² This is a governmental unit or tract measure of size (MOS), the sum of estimated occupied housing units in the governmental unit or tract.

Table 3. Five Year CVs for Typical Tracts by Size Class based on Sample Simulations

Tract Size Category (Tract MOS)	Total Tracts	Average Tract Size	Current Stratification		New Stratification	
			CV 2.9 Million	CV 3.54 Million	CV 2.9 Million	CV 3.54 Million
Total:	64,400					
0 - 400	1,051	291	66%	59%	41%	35%
401 - 1,000	9,990	766	41%	36%	30%	25%
1,001 - 2,000	31,011	1,485	29%	26%	29%	25%
2,001 - 4,000	19,893	2,636	25%	23%	29%	25%
4,001 - 6,000	1,889	4,684	19%	17%	29%	25%
6,001 +	566	8,337	15%	13%	28%	25%

Notes and assumptions:

1. CVs are based on simulated samples.
2. Data is based on Supplemental 2010 valid housing unit address counts at the tract level.
3. A 10 percent population characteristic is used ($p = 0.1$).
4. An ACS design factor of 2.1 is used (average of three poverty design factors).
5. A block level occupancy rate (average of ACS and Census 2000 rates) and a 2.6 persons per household is used.
6. A 100 percent computer assisted personal interview (CAPI) interview rate is assumed.
7. The lowest tract size category does not include tracts with MOS < 150.
8. Ungeocoded addresses cannot be included.

Table 4 shows the universe size and the sample sizes generated by each of the four simulations. Note that this table excludes the addresses in the 1,043 tracts with an estimated number of occupied housing units less than 150 as well as all ungeocoded addresses in the 2010 supplemental U.S. edited Master Address File extract.

Table 4. Universe and Simulated Sample Sizes by Tract Size Category

Tract Size Category (Tract MOS)	Total Tracts	Average Tract Size	Universe Size	Current Stratification		New Stratification	
				Sample 2.9 Million	Sample 3.54 Million	Sample 2.9 Million	Sample 3.54 Million
Total	64,400		132,636,915	2,837,551	3,463,324	2,859,978	3,488,879
0 - 400	1,051	291	377,586	11,330	13,573	21,127	25,372
401 - 1,000	9,990	766	8,897,949	253,427	306,356	391,918	473,556
1,001 - 2,000	31,011	1,485	50,996,796	1,299,697	1,583,020	1,426,162	1,729,380
2,001 - 4,000	19,893	2,636	57,363,023	1,027,744	1,259,093	906,799	1,117,791
4,001 - 6,000	1,889	4,684	9,781,981	161,550	198,358	87,250	109,219
6,001 +	566	8,337	5,219,630	83,803	102,924	26,722	33,561

To demonstrate the potential effect the new strata may have on total variance at the county level through additional weight variation, Table 5 shows the distribution of the square root of the ratio (R) of the sum of the squared weights from the simulated sample under the research stratification to the simulated sample under the current stratification, each at the 3.54 million level. This distribution shows that the majority (2,928) counties have ratios between 0.8 and 1.2. Clearly the new stratification has little effect at the county level. 204 counties do show that the variance will increase, though not appreciably. We surmise that these counties are most likely large, since they must contain multiple sampling strata. If so, the

standard errors are most likely small. Table 6 shows four selected large counties typical of other counties with a large value of R. From AmericanFactfinder, for “number of households in poverty”, we see that indeed the standard errors (CVs) were relatively small to begin with. As such, even moderate percent increases in variance due to increased weight variation still produce relatively small CVs. A case in point is Gwinnett County, GA. Even with an approximate 47% increase in the CV, when applied to the already relatively small CV of 3.83% yields a small CV of 5.63%, which is still a very reliable estimate. This is typical of the results we observed in all large counties and may be generalized to all ACS estimates.

Table 5. Distribution of County Level Ratios – New to Current Stratification (3.54 Million Sample)

County Size (Valid Addresses)	$0.6 \leq R^{(1/2)} < 0.8$	$0.8 \leq R^{(1/2)} < 1$	$1 \leq R^{(1/2)} < 1.2$	$1.2 \leq R^{(1/2)} < 1.4$	$1.4 \leq R^{(1/2)} < 1.6$	$1.6 \leq R^{(1/2)} < 1.8$
Totals	11	1,218	1,710	156	42	6
0 - 1K	0	52	12	0	0	0
1K - 2K	2	106	36	0	0	0
2K - 3K	7	126	22	0	0	0
3K - 4K	1	121	38	0	0	0
4K - 5K	1	116	53	0	0	0
5K - 6K	0	104	57	1	0	0
6K - 7K	0	68	46	2	0	0
7K - 8K	0	64	69	3	0	0
8K - 9K	0	58	53	1	0	0
9K - 10K	0	52	57	1	0	0
10K - 11K	0	33	59	4	0	0
11K - 12K	0	29	59	5	0	0
12K - 13K	0	24	49	3	0	1
13K - 14K	0	23	52	4	0	0
14K - 15K	0	17	37	2	1	0
15K - 20K	0	79	196	7	4	1
20K - 25K	0	43	134	10	3	0
25K - 100K	0	78	483	66	22	4
100K - 500K	0	22	176	40	12	0
500K+	0	5	22	7	0	0

County Level Ratio = $[\text{New Design } \sum_{\text{sample}} (w^i)] + [\text{Current Design } \sum_{\text{sample}} (w^i)]$, $w = (1/\text{rate}_{\text{blk}})$

Table 6. Estimated Change in CVs for Select Large Counties

County, State	3 yr ACS Data - American FactFinder			CV	$R^{(1/2)}$	Estimated New CV
	Households	Households in Poverty	M.O.E.			
Gwinnett County, Georgia	256,260	19,719	1,250	3.83%	1.4685	5.63%
Denton County, Texas	203,429	14,034	1,182	5.09%	1.4992	7.63%
Bronx County, New York	471,553	127,816	3,094	1.45%	1.0356	1.52%
Bell County, Texas	95,646	10,781	930	5.23%	1.2299	6.41%

Recently, we conducted an in depth analysis of the current sample design by calculating and analyzing coefficients of variation (CVs) for a 10 percent estimate of the person poverty rate.

The following questions were addressed in this research:

- *Can the stratification scheme for the ACS be modified in a way which allows for a more equitable distribution of sample minimizing the differences of the CVs across areas?*
- *What effects would the projected sample increase from 2.9 million to 3.54 million have on a stratification redesign?*

The goal of the research was to minimize the differences of the coefficients of variation for tract level estimates by size class. We used a 10 percent person poverty rate estimate. Based on our results, the new stratification will use sixteen sampling strata, each with a corresponding sampling rate. Three of these rates, for the smallest governmental units, will be fixed and 13 will be functions of the base rate. The sixteen strata and their defined sampling rates are shown in Table 2.

Table 3 shows a comparison of CVs for the current and new stratification, using both the current and expanded sample size, respectively.

Table 2. New Stratification and Annual Sampling Rates

Stratum	Block MOS Criteria	Annual Sampling Rate
1	800 ≤ GUMOS ≤ 1200	2.8 x Base Rate
2	400 ≤ GUMOS < 800	Fixed
3	200 ≤ GUMOS < 400	Fixed
4	0 ≤ GUMOS < 200	Fixed
5	0 ≤ TRACTMOS < 400	3.5 x Base Rate
6	0 ≤ TRACTMOS < 400 High Response	0.92 x 3.5 x Base Rate
7	400 ≤ TRACTMOS < 1000	2.8 x Base Rate
8	400 ≤ TRACTMOS < 1000 High Response	0.92 x 2.8 x Base Rate
9	1000 ≤ TRACTMOS < 2000	1.7 x Base Rate
10	1000 ≤ TRACTMOS < 2000 High Response	0.92 x 1.7 x Base Rate
11	2000 ≤ TRACTMOS < 4000	Base Rate
12	2000 ≤ TRACTMOS < 4000 High Response	0.92 x Base Rate
13	4000 ≤ TRACTMOS < 6000	0.6 x Base Rate
14	4000 ≤ TRACTMOS < 6000 High Response	0.92 x 0.6 x Base Rate
15	6000 ≤ TRACTMOS	0.35 x Base Rate
16	6000 ≤ TRACTMOS High Response	0.92 x 0.35 x Base Rate