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

June 11, 2008

2007 AMERICAN COMMUNITY SURVEY RESEARCH MEMORANDUM SERIES #ACS07-R-10

MEMORANDUM FOR Susan Schechter
Chief, American Community Survey Office

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

Prepared by: Edward C. Castro Jr. and Steven P. Hefter
American Community Survey Sample Design Branch
Decennial Statistical Studies Division

Subject: American Community Survey: Analysis of the Effects on the Sample Distribution of Using Minor Civil Divisions as Design Areas

I. Introduction

The purpose of this memorandum is to assess the impact on the distribution of the American Community Survey (ACS) sample by including Minor Civil Divisions (MCDs) in the twelve "strong" MCD states as design areas when determining the block level sampling strata assignment – and thus the sampling rate – for the ACS[1]. Beginning with the 2002 sample selection, MCDs were considered for inclusion as design areas. This decision was made for two reasons: 1) MCDs in the 12 "strong" MCD states were included in the Census 2000 Long Form sample design; and 2) The desire to produce reliable ACS estimates for MCDs in the 12 "strong" MCD states[2].

II. Background

The ACS selects a sample of housing unit addresses from the Master Address File (MAF) twice a year. The Main sampling phase occurs in August/September of the year previous to the sample year and accounts for 99 percent of the sample. In January of the sample year, a sample of addresses that have been added to the MAF since the Main MAF extracts were created is selected. This is known as the Supplemental sampling phase and accounts for approximately one percent of the total ACS sample.

There are two steps, or stages, in each sampling phase: first-stage and second-stage sampling. The first-stage sample comprises approximately 20 percent of the addresses. The other 80 percent is allocated to four backsamples, ensuring no MAF housing unit

[1] The "strong" MCD states are those states that have functioning governmental entities at the sub-county level.
[2] Note that Puerto Rico contains no MCDs and therefore is not included in the research.

address record is eligible for sampling more than once in any five-year period. The second-stage sample is selected from the current year's first-stage sample for data collection in the sample year. The second-stage sample is selected based upon which second-stage sampling stratum each block is in.

The current ACS sample design assigns each block (in current geography) to one of five second-stage sampling strata based on a measure of size (MOS) calculated for each design area which includes all or part of the block. This is done during the Main sampling phase. The set of design areas considered are:

- Counties, County Equivalents, and Municipios in Puerto Rico
- Places
- School Districts – *elementary, secondary, and unified*
- Minor Civil Divisions in the 12 "strong" MCD states (Connecticut, Maine, Massachusetts, Michigan, Minnesota, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont, Wisconsin)
- American Indian areas
- Tribal Subdivisions
- Alaska Native Village Statistical Areas
- Hawaiian Homelands

Note: Only design areas that are flagged as active on the Geographic Reference File – Names are used. All American Indian areas, Alaska Native Village Statistical Areas, Hawaiian Homelands, School Districts, and counties are considered active.

Each block may be in several different design areas, each with its own MOS. The smallest MOS (of the set of applicable MOSs for each block) is determined and is referred to as the Governmental Unit MOS (GUMOS)[3]. The MOS for each Census Tract is also determined, assigned to each block as appropriate, and is referred to as HUTRACT. At this point, each block in the U.S. and Puerto Rico has two MOSs associated with it, GUMOS and HUTRACT. These two measures are used to assign each block to a second-stage sampling stratum (SBSTR) using the following algorithm:

If (0 < GUMOS < 200) then SBSTR='5'
Else If (200 ≤GUMOS < 800) then SBSTR='2'
Else If (800 ≤GUMOS ≤1200) then SBSTR='3'
Else If (HUTRACT ≥2000) then SBSTR='4'
Else SBSTR='1'

The block level sampling rate for each stratum is determined by first calculating a base sampling rate. The sampling rate for four of the five sampling strata is then calculated based upon the base sampling rate. The base sampling rate calculation incorporates

[3] For further details see 2007 ACS Sampling Memorandum Series #ACS07-S-1, "*Creating the Governmental Unit Measure of Size (GUMOS) Datasets for the American Community Survey and the Puerto Rico Community Survey*" dated June 6, 2006.

projected growth between the Main and Supplemental phases to yield an annual sample size of approximately 3 million addresses.

The base sampling rate (BR) is rounded to four decimal places and is the smallest number such that:

$0 \leq BR \leq 1$, and;

$$\sum_{SBSTR=2} 3 \times BR + \sum_{SBSTR=3} 1.5 \times BR + \sum_{SBSTR=1} BR + \sum_{SBSTR=4} 0.735 \times BR + \sum_{SBSTR=5} 0.10 \geq \left(3{,}000{,}000 - \sum_{\text{all SBSTR}} \text{projected growth} \right)$$

where the index on the summation runs through all valid addresses in the SBSTR.

Table 1 shows the relationship between the sampling rates and the base rate.

Table 1. Sampling Rate Definitions

Sampling Stratum	Target Sampling Rate
'1'	Base Rate
'2'	3 x Base Rate
'3'	1.5 x Base Rate
'4'	0.735 x Base Rate
'5'	0.10

The sampling rates in strata '1' and '4' are then modified by a reduction factor for blocks in tracts with high expected mail and Computer Assisted Telephone Interview (CATI) cooperation rates. This is to offset the cost of the differential Computer Assisted Personal Interview (CAPI) sampling in areas with low cooperation that are sampled at higher rates[4].

III. Simulation Methodology

MCDs in the 12 "strong" MCD states were implemented as design areas beginning in 2002. Several other changes were also implemented in 2002 including changing the sampling rates in many counties. These additional changes to the sample design make it difficult to isolate the effect of adding MCDs as design areas relative to the 2002 sample. In order to isolate the effect that MCDs have on the geographic distribution of the

[4] For further details see 2007 ACS Sampling Memorandum Series #ACS07-S-3 "*Specifications for Selecting the Main and Supplemental Housing Unit Address Samples for the American Community Survey*" dated August 23, 2006.

sample, the 2007 sample selection has been simulated two ways: with MCDs (referred to as the production simulation) and without considering MCDs (referred to as the research simulation) as design areas.

The following steps were taken to produce simulated Main 2007 sample sizes under both scenarios. The actual Main 2007 sample selection was simulated instead of using the actual sample counts so that valid comparisons of the results could be made.

A. Calculate the New Measures of Size

GUMOS was determined for each block in the nation. In addition, a tract measure of size, HUTRACT, was also calculated for each current tract in the nation.

B. Assign Blocks to a Second-Stage Sampling Stratum

Based upon the algorithm given in Section II, each current block in the nation was assigned to one of the five second-stage sampling strata.

C. Calculate the Base Sampling Rate

The number of valid addresses within each second-stage sampling stratum is shown in Attachment A. Using these totals together with the projected growth from the Main phase to the Supplemental phase, the base sampling rate was calculated.

D. Determine the Sampling Rate

The sampling rate for each stratum was calculated using the formulae given in Table 1. The differential CAPI sampling reduction factor was applied to blocks that were both in sampling strata '1' and '4' and in tracts with the highest expected mail/CATI cooperation rates.

E. Calculate Block Level Sampling Rates

Using the sampling rate determined in the previous step, the expected sample size was determined for each block. This was done by multiplying the sampling rate by the number valid housing unit addresses in each block.

F. Create Estimated Sample Sizes for Each State

The block level sample estimates were summed to the state level and rounded to the nearest integer. They are shown in Attachments B and C.

IV. Results

We would expect there to be an increase in the expected sample size in the 12 "strong" states by including MCDs as design areas. Attachment B shows the simulated sample totals for the 12 "strong" MCD states. Due to the fixed target sample size and the increase in sample in these states, the sample must therefore decrease in the rest of the country. Due to the fact that the non-strong MCD states do not use MCDs as sampling entities, the second-stage sampling stratum assignment wasn't affected by their removal. Attachment C clearly shows that for the rest of the country, the sample decreased. This is due to the smaller base sampling rate (from 0.0231 for the research simulation to 0.0223 for the production simulation).

Unexpectedly, there were four "strong" MCD states where using MCDs as design areas caused the sample to decrease. New Jersey saw the largest such drop, where the sample decreased from 74,074 to 71,593. This is due to the fact that most of the MCDs in these states have an MOS greater than 1,200. This means that adding MCDs as sampling entities did not increase the sampling rate, since all blocks being assigned to a second-stage sampling stratum used either another MOS less than 1,200 to determine GUMOS, which did not change, or HUTRACT, which also did not change, was being used to assign the sampling stratum in the first place. For example, Rhode Island has 31 MCDs, but only one had an MOS less than 1,200. In the 30 MCDs with an MOS greater than 1,200, GUMOS was also greater than 1,200 and HUTRACT was being used to assign the blocks to a second-stage sampling stratum. HUTRACT did not change between simulations and the assignment of the second-stage sampling stratum remained unchanged. Thus only the smaller base sampling rate impacted the sample size in these four states.

Additionally, there were four "strong" MCD states (Michigan, Minnesota, Pennsylvania, and Wisconsin) where the sample increased by about 20,000 addresses. This is due to the shift of addresses in these states to the second-stage sampling strata with higher sampling rates. These four states drive the total increase in the sample size of the 12 "strong" MCD states of almost 87,000 addresses.

All non-strong MCD states, as expected, showed a decrease in the sample size because any increase to the sample in the "strong" states must be offset by a corresponding decrease in the balance of the states. The total impact of this was a decrease in the sample in these states of about 70,000 addresses.

The sample for the nation as a whole increased from 2,844,625 under the research simulation to 2,860,718 addresses under the production simulation. By using MCDs as sampling entities, a portion of the housing unit address inventory was shifted out of the second-stage sampling strata where sample was reduced for CAPI (SBSTR '1' and '4'). Much of this sample was moved into the second-stage sampling strata of the smallest areas (SBSTR '2' and '5'), accounting for a 33 percent increase in sample in both strata,

100,000 in SBSTR ‘2’ and 30,000 in SBSTR ‘5’. This can be seen in Attachment A. Thus, the total sample size increased by including MCDs in the GUMOS calculation.

V. Conclusions

Using MCDs as design areas impacts all states in the nation. It causes an increase to the sample in some states while offsetting that with a decrease in the balance of the nation. The main effects of including MCDs as design areas are:

- A larger sample in most “strong” MCD states;
- A smaller sample in some “strong” MCD states due to large MCD MOS;
- A corresponding smaller sample in rest of nation; and
- A relatively small increase in the sample for the nation as a whole due to a change in the second-stage sampling stratum distribution, where addresses were shifted out of the CAPI reduction strata.

VI. Contact

Please contact Steven Hefter at (301) 763-4082 or Edward Castro at (301) 763-3427 with comments or questions.

cc: A. Navarro (DSSD)
K. King
M. Asiala
S. Hefter
A. Williams
M. Beaghen
K. Albright
D. Keathley
J. Powers
R. Sirkis

Attachment A

Valid Address and Expected Sample Distribution by Second-Stage Sampling Stratum for Each Simulation

Without MCDs as Sampling Entities (Research Simulation)

Second-Stage Sampling Stratum	Sampling Rate	Valid Addresses	Simulated Sample Size
1	Base Rate*	63,542,087	1,398,354
2	3 x Base Rate*	4,318,509	299,273
3	1.5 x Base Rate*	2,683,863	92,996
4	0.735 x Base Rate*	60,278,055	964,047
5	10%	899,552	89,955
Total		131,722,066	2,844,625

* Base Rate = 0.0231

With MCDs as Sampling Entities (Production Simulation)

Second-Stage Sampling Stratum	Sampling Rate	Valid Addresses	Simulated Sample Size
1	Base Rate*	61,595,681	1,309,498
2	3 x Base Rate*	5,968,462	399,290
3	1.5 x Base Rate*	3,335,927	111,587
4	0.735 x Base Rate*	59,625,974	920,741
5	10%	1,196,022	119,602
Total		131,722,066	2,860,718

* Base Rate = 0.0223

Attachment B

Expected Sample Sizes for 12 "Strong" MCD States

State	Research Simulation	Production Simulation	Difference
Connecticut	29,159	28,288	-871
Maine	19,740	23,967	4,227
Massachusetts	52,977	52,366	-611
Michigan	**103,541**	**121,621**	**18,080**
Minnesota	**56,857**	**77,367**	**20,510**
New Hampshire	13,838	14,910	1,072
New Jersey	74,074	71,593	-2,481
New York	174,641	179,573	4,932
Pennsylvania	**122,323**	**142,374**	**20,050**
Rhode Island	8,936	8,627	-309
Vermont	12,037	12,120	83
Wisconsin	**59,263**	**81,448**	**22,185**
Total	727,385	814,251	86,867

Attachment C

Expected Sample Sizes for 38 Non-Strong MCD States Plus The District of Columbia

State	Research Simulation	Production Simulation	Difference
Alabama	52,443	50,691	-1,752
Alaska	9,964	9,698	-266
Arizona	55,701	53,800	-1,901
Arkansas	31,850	30,836	-1,014
California	272,386	263,107	-9,279
Colorado	46,276	44,711	-1,565
Delaware	10,400	10,056	-344
District of Columbia	6,013	5,805	-208
Florida	163,865	158,207	-5,658
Georgia	81,059	78,322	-2,737
Hawaii	12,265	11,853	-412
Idaho	15,513	15,012	-501
Illinois	120,607	116,583	-4,024
Indiana	61,984	59,897	-2,086
Iowa	39,465	38,258	-1,207
Kansas	33,031	32,002	-1,030
Kentucky	43,036	41,594	-1,442
Louisiana	47,936	46,315	-1,621
Maryland	46,959	45,349	-1,611
Mississippi	29,041	28,076	-965
Missouri	66,103	63,978	-2,124
Montana	14,582	14,157	-425
Nebraska	25,434	24,709	-725
Nevada	22,079	21,323	-755
New Mexico	21,522	20,817	-705
North Carolina	85,123	82,260	-2,863
North Dakota	11,789	11,463	-326
Ohio	112,358	108,575	-3,782
Oklahoma	47,854	46,328	-1,526
Oregon	34,753	33,581	-1,172
South Carolina	42,699	41,252	-1,446
South Dakota	11,898	11,548	-350
Tennessee	57,112	55,158	-1,954
Texas	211,210	204,038	-7,172
Utah	21,496	20,796	-700
Virginia	63,695	61,516	-2,179
Washington	60,040	58,002	-2,038
West Virginia	21,449	20,738	-711
Wyoming	6,252	6,056	-197
Total	**2,117,241**	**2,046,467**	**-70,774**