

**AMERICAN COMMUNITY SURVEY:
SAMPLE DESIGN FOR COMPUTER ASSISTED PERSONAL INTERVIEW**

Mark E. Asiala
Decennial Statistical Studies Division
U.S. Census Bureau

For presentation at the April 25-27, 2005 Meetings of the Census Advisory Committee on the African American Population, the American Indian and Alaska Native Populations, the Asian Population, the Hispanic Population, and the Native Hawaiian and Other Pacific Islander Populations

What was the purpose of this research?

The purpose of this research was to devise a methodology to reduce the disparity in the reliability of tract-level estimates from the American Community Survey (ACS).

What was the special interest in the reliability of estimates for tracts?

In general, tracts are small geographic areas with small populations. This makes estimates for these areas more susceptible to variations in the interviewed sample which, in turn, causes greater disparity in the reliability of estimates produced for those tracts.

While all surveys are affected by noninterviews which directly affects the interviewed sample size, the design of the ACS adds another source of variation.

The ACS design has three modes of data collection, mail, Computer Assisted Telephone Interviewing (CATI), and Computer Assisted Personal Interviewing (CAPI). Because of the higher relative cost of personal interviewing as compared to mailing forms and conducting telephone interviews, a 1-in-3 sub-sample is taken of those whose response is not received by mail or CATI to follow up in the CAPI mode.

Tracts which have a higher proportion of their sample sent to CAPI will be impacted more greatly by the 1-in-3 sub-sampling performed in the CAPI operation. The result is that fewer interviews are obtained for the tract which leads to decreased reliability of estimates produced for that tract.

What was the scope of this research?

The scope of this research was to identify tracts of low combined mail and CATI cooperation rates and to designate an increased CAPI sub-sampling rate for these tracts. For our research, the cooperation rate is defined as the ratio of the mail interviews plus the CATI interviews divided by the estimated occupied mailable universe. By following up on a larger proportion of the nonrespondents from mail and CATI, more interviews can be obtained for these tracts and the reliability of estimates produced for these tracts can be improved.

For this research, only tracts that were at least 75 percent mailable, as defined by the ACS, were considered to be in scope for having their CAPI sub-sampling rate increased. The rationale for this decision was that tracts which have high unmailable rates (defined as having 25 percent or more of the addresses classified as unmailable) already benefit from a 2-in-3 sub-sampling in the CAPI phase for all unmailable addresses. Thus, there are mechanisms already in place to improve the interviewed sample size for this class of tracts. This research thus focused on tracts which are mostly mailable but for which we have observed a low cooperation rate through either the ACS or the Census 2000 long form.

What were the constraints?

The implementation of the differential sub-sampling operation was designed to be cost neutral. Thus, any additional expense that may be incurred from increasing the CAPI sample rate in certain tracts must be offset by a reduction in expenses elsewhere so that the total ACS cost remains unchanged.

How were tracts of low cooperation rate identified?

Based on Census 2000 long-form mail return rates and historical ACS mail and CATI data, a projected combined mail and CATI cooperation rate was calculated for each tract where data existed. The cooperation rate was then used to classify tracts into categories of low, medium, high, and highest cooperation rates.

How was the CAPI sub-sampling rate changed?

Tracts in the low cooperation rate group had their CAPI sub-sampling rate increased from a 1-in-3 to a 1-in-2 rate. The subsampling rate for tracts in the medium cooperation rate group was increased from a 1-in-3 to the a 2-in-5 rate. All other tracts remained at the 1-in-3 CAPI sub-sampling rate. The objective of this design change was to reduce the difference in the percentage of interviewed cases at the census tract level. Thus, the disparity in the reliability of tract level estimates should be reduced.

Four options were researched which used different cutoffs to define the cooperation rate categories. Of these options, Option #2 was chosen based on achieving the objective of reducing the disparity of the reliability of the tract level estimates for the four cooperation rate categories. Table 1 provides the final projected cooperation rate categories and their corresponding CAPI sampling rate.

Table 1: Classification of Tracts based on Projected Cooperation Rate (Option #2)

Projected Cooperation Rate(%)	CAPI Sampling Rate
0-35	1-in-2
36-50	2-in-5
51-60	1-in-3
61-100	1-in-3

How were the costs offset?

To make this change cost neutral, we reduced the initial sample by up to eight percent in those tracts with a projected cooperation rate of 61% or higher. Portions of tracts that are being sampled at a higher rate because of the size of the smallest governmental unit in which it is contained were not reduced. This reduction in sample offsets the increased

cost in the CAPI component of data collection that is due to the increase in the CAPI sub-sampling rate.

Note that tracts that are not in scope for this plan due to high unmailable rates did not have their sample reduced regardless of their cooperation rate. Also, tracts in the high cooperation rate category, whose cooperation rate is in the 51–60 percent range, did not have their sample reduced and will continue with the 1-in-3 sub-sampling rate.

How many tracts were affected?

There are approximately 65,200 tracts in the nation. As a result of this plan, 12,191 tracts had their CAPI sub-sampling rate increased. The initial sample for 42,316 tracts was reduced. The balance of tracts, including those that are out of scope, were not affected by this plan. The table below summarizes the number of tracts by the CAPI sampling rate and the application of sample reduction.

Table 2: Distribution of Tracts by CAPI Sampling Rate and Reduction in Sample

CAPI Sampling Rate	Reduction in Sample	Number of Tracts
1-in-2	No	4,420
2-in-5	No	7,771
1-in-3	No	10,661
1-in-3	Yes	42,316

What is the expected benefit in reliability of tract-level estimates from this design change?

The benefit from this change is concentrated in the tracts designated for the increased CAPI sub-sampling rates. The measure used for estimating the reliability in these tracts is the coefficient of variation or CV. The CV is defined as the standard error divided by the estimate. CVs are calculated for a theoretical 5-year estimate of a 10 percent population characteristic. The 1-in-2 sub-sampled tracts show an improvement as measured by the mean CV across tracts in this group of about 4.2 percentage points, dropping from 26.44 to 22.25. The 2-in-5 sub-sampled tracts show an improvement in the CV from 21.96 to 20.46 percent. The tracts which are subject to the sample reduction as part of the offsetting cost procedure show only a small increase in the CV of 0.7 percentage points, rising from 16.84 to 17.53 percent. The comparison of the Original design to the New design is presented in Table 3 below.

Table 3: Tract-level Mean, Median, and Standard Deviation of CVs by CAPI Sampling Rate

		Tract-Level CVs			
CAPI Sampling Rate	Sample Design	Number			Std.
		Tracts	Median	Mean	Dev.
All	Original Design	63,042	17.27	18.51	5.91
	New Design	63,056	17.51	18.50	5.37
1-in-2	Original Design	4,330	23.81	26.44	9.51
	New Design	4,349	19.75	22.25	8.60
2-in-5	Original Design	7,749	20.38	21.96	6.22
	New Design	7,749	18.90	20.46	6.03
1-in-3 (no reduction)	Original Design	8,718	18.50	19.61	4.87
	New Design	8,718	18.50	19.61	4.87
1-in-3 (with reduction)	Original Design	42,245	16.23	16.84	4.38
	New Design	42,240	16.93	17.53	4.54

Notes: (1) Original Design has all tracts with a 1-in-3 CAPI sample.

(2) The number of tracts is less than 65,200 because a minimum interviewed sample size of 10 was used to screen out small tracts for these calculations.

What will be the benefit from a data user perspective?

Estimates produced for tracts identified by this plan for increased CAPI sub-sampling will be more reliable and the 5-year estimates will be more stable. While the research and implementation of this plan was to reduce disparity in reliability of tract level estimates, the implementation of this plan will increase interviewed sample for certain demographic groups. To the degree that those tracts have higher concentrations of certain demographic groups, estimates concerning those groups should also be of higher quality.

Census 2000 data in tables 4 and 5 below shows that certain racial and ethnic demographic groups are more concentrated in the 1-in-2 and 2-in-5 sub-sampled categories than in the nation as a whole (the “All” group). If these demographic distributions hold for 2005, we can expect that a higher number of non-Whites and Hispanics will be interviewed in our national sample under the new design. Blacks and Hispanics are two notable groups where the concentration in the 1-in-2 designated tracts is three times the national average. Based on Census 2000 data, these same tables show that American Indians are more concentrated in the out of scope tracts. **This is because of the high unmailable rate for many of the American Indian reservations.** A greater proportion of census tracts in American Indian areas, however, will be sampled at the higher CAPI sub-sampling rate of 2-in-3 used for unmailable addresses. All other non-White groups shows the potential for an increase in interviewed sample cases.

Table 4: Racial Distribution Within Group

Group	Whites	Blacks	American Indian	Asian	Native Hawaiian	Some Other Race	Two or More Races	Total
All	75.2%	12.3%	0.9%	3.6%	0.1%	5.5%	2.4%	100.0%
1-in-2	35.2%	36.3%	0.9%	4.5%	0.2%	18.7%	4.3%	100.0%
2-in-5	50.8%	25.5%	0.9%	5.5%	0.3%	13.1%	3.9%	100.0%
1-in-3 (no reduct.)	66.1%	17.1%	0.9%	4.9%	0.2%	7.7%	3.2%	100.0%
1-in-3 (w/ reduct.)	84.5%	7.2%	0.6%	3.1%	0.1%	2.6%	1.9%	100.0%
Out of Scope	78.1%	6.1%	9.1%	0.9%	0.2%	3.4%	2.1%	100.0%

Note: This table reflects the racial distribution using Census 2000 data. The current racial distribution may have changed since 2000.

Table 5: Hispanic Distribution Within Group

Group	Non-Hispanic	Hispanic	Total
All	87.5%	12.5%	100.0%
1-in-2	63.0%	37.0%	100.0%
2-in-5	71.9%	28.1%	100.0%
1-in-3 (no reduct.)	82.1%	17.9%	100.0%
1-in-3 (w/ reduct.)	93.3%	6.7%	100.0%
Out of Scope	89.2%	10.8%	100.0%

Note: This table reflects the ethnic distribution using Census 2000 data. The current ethnic distribution may have changed since 2000.

When was this implemented?

The new design based on this research was implemented for the 2005 ACS full implementation that started in January 2005.