Redesigning the American Community Survey Computer Assisted Personal Interview Sample

Edward C. Castro Jr., Steven P. Hefter
U.S. Census Bureau
Washington, DC 20233

Abstract

The American Community Survey (ACS) collects data in three phases: mailout/mailback, Computer Assisted Telephone Interview (CATI), and Computer Assisted Personal Interview (CAPI). During the CAPI phase, a sample of mail/telephone non-respondents and addresses deemed to be unmailable is selected for a personal visit. The sampling rate for this process is determined by the CAPI sampling stratum for the tract in which the address resides, which is assigned based upon the mail/telephone cooperation rate of the tract. The initial assignment was made prior to 2005. Based upon recent ACS data, cooperation rates were re-calculated for all tracts, which were then assigned to new CAPI sampling strata. This paper discusses the data that led us to revisit the CAPI sampling stratum assignments, and provides the methodology used to calculate the tract level mail/telephone cooperation rates used to assign each tract to a CAPI sampling stratum.

Keywords: ACS, Sampling, Nonresponse, Personal Interview

1. Introduction

The ACS is a rolling monthly survey that collects socio-economic and demographic data describing the population and housing inventory in the United States similar to what was historically collected on the census long form. From 2000 to 2005, 1,239 counties were included in the Census 2000 Supplemental Survey (C2SS), with the addition of Broomfield County, Colorado in 2002. In 2005, the sample was expanded to include all 3,141 counties and county equivalents in the United States. All 78 municipios in Puerto Rico are included in a separate survey, the Puerto Rico Community Survey, which uses the same basic design as the ACS.

2. Sample Design

The ACS selects a sample of housing unit addresses from the Master Address File (MAF) twice a year [2]. Main sampling occurs in August/September of the year prior to the sample year. Approximately 99 percent of the total annual ACS sample is selected at this time. 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 Supplemental sampling and accounts for approximately one percent of the total annual ACS sample [5].

The ACS employs three modes of data collection: mail, CATI, and CAPI. In general, questionnaires not completed during one phase are sent to the next phase of data collection. The exception to this is for unmailable addresses, which are sampled and sent directly to the CAPI mode. A sample of the mail/CATI non-responding addresses is also sent to CAPI for follow-up.

Prior to March 2005 all addresses that did not respond to the mailout questionnaire and that were not reached during the CATI phase of data collection were sampled for CAPI at a flat 1-in-3 rate. In order to produce estimates of comparable reliability for all areas, we developed methodology to differentially sample non-responding addresses for CAPI at the tract level by increasing the CAPI sampling rates in tracts with low combined mail/CATI cooperation rates. To offset the additional cost associated with increasing the number of CAPI cases, the differential CAPI sampling plan was designed to be cost neutral. This was accomplished by reducing the initially selected sample in blocks with the two lowest sampling rates in tracts with the highest expected mail/CATI cooperation rates. The sample reduction in these areas began with the Main 2005 sample selection in August-September of 2004 and has been reflected in the mailout beginning with the January 2005 panel. Note that the CAPI sampling rate for all unmailable addresses was not changed and remains at 2-in-3 [4].

3. Research Question

Our research was designed to answer the following questions:

Does the allocation of tracts to differential CAPI sampling strata need to be revised? If so, how will this be accomplished?

The answer to the overarching research question rested on several secondary questions:

1. *Does the cooperation rate for each individual stratum fall within the predicted range for that stratum at various levels of geography?* This will enable us to assess how well the current design is working overall.

2. *What is the distribution of the tract level cooperation rates by assigned CAPI sampling strata?* This will assess how well the current design is working for each tract.

3. *How should the vacancy rate component of the cooperation rate be calculated?* The vacancy rate has a significant impact on the cooperation rate calculation. Any change to the vacancy rate calculation will have an impact on the design.

4. *What impact is there on the CAPI workload from a new allocation of tracts to the CAPI sampling strata?* Any new allocation of tracts into new CAPI strata will impact the CAPI workloads and could necessitate a change to the CAPI reduction factor as the redesign was also constrained to be cost neutral.

In order to address these questions, indicators related to cooperation and return rates, expected sample size, and cost estimates for small levels of geography were created. We classified each tract into a CAPI sampling stratum. The cooperation rates, which remove an estimate of the number of vacant and deleted units from the denominator, were used to group tracts by level of cooperation. The return rates from the mail, CATI, and CAPI phases of data collection were used in conjunction with the estimated sample sizes and cost estimates to determine a new value of the reduction factor.

4. Previous CAPI Sample Design

4.1 Combined Mail/CATI Cooperation Rate Calculation

We excluded vacant housing units from the denominator of the cooperation rate so that a high vacancy rate alone will not cause a tract to be sampled at a higher rate. The final estimate of the cooperation rate was defined as the weighted total of mail and CATI interviews divided by the estimated number of occupied mailable addresses. We calculated the cooperation rates using C2SS data from 2000 through 2003. Thus, they could only be calculated for about 50,000 tracts in the nation. For the approximately 15,000 tracts with no ACS data, cooperation rates were modeled using a simple linear model from the Census 2000 long-form cooperation rates [1].

4.2 CAPI Stratum Design

The initial research analyzed several design options. There were two important factors considered: 1) What cooperation rate cut-offs were appropriate to use in assigning the CAPI sampling strata and 2) What reduction factor should be used in the high cooperation rate strata to keep the sample design cost neutral. Table 1 shows the cutoff values in the final design.

Table 1. CAPI Strata Cut-offs

CAPI Sampling Stratum	Cooperation Rate	Sampling Rate
1	0% - 35%	1-in-2
2	36% - 50%	2-in-5
3	51% - 60%	1-in-3
4	61% - 100%	1-in-3 (with reduction)

These cut-offs values were chosen for several reasons. They help increase the overall interview rate for groups with typically low response patterns. This increase in the number of interviews in turn reduces the variance of the published estimates. In addition, these cut-offs provided a good balance between the number of tracts where the sampling rate would increase and maintaining a reasonable reduction factor. The CAPI sample was increased in approximately 11,000 tracts while the original sample was reduced by 8% in roughly 42,000 tracts. These are the tracts that are in CAPI sampling stratum 4 and contain blocks with high population density. Tracts in CAPI sampling stratum 4 but in areas of low population density do not undergo sample reduction. Thus, a significant number of areas were impacted by increasing the sampling rate while spreading out the compensating sample reduction.

5. Decisions, Assumptions, and Limitations

5.1 Decisions and Assumptions

The following questions needed to be answered prior to completing the research.

1. *Should an estimate of deleted units be removed from the denominator of the cooperation rate?*

In the initial research, the cooperation rate was calculated by taking the total weighted number of mail and CATI interviews together divided by the estimated number of occupied mailable addresses. The number of occupied mailable addresses was estimated by

subtracting the estimated number of vacant mailable addresses from an estimate of the number of mailable addresses. In this way, a high vacancy rate would not cause an artificially low cooperation rate. The same logic can be applied to deleted units as these should not "hurt" a tract's cooperation rate. We decided to remove an estimate of the deleted units from the denominator of the cooperation rate.

2. ***Should the CAPI completion rate be taken into account since the overall goal is to produce estimates of comparable reliability across all areas?***

The methodology used in the assignment of the differential CAPI sampling strata was based entirely upon the mail/CATI cooperation rate. If there was a large change for certain types of areas in the CAPI completion rate, then it might be possible to use this rate to better determine the CAPI stratum assignment. A potential problem with using the CAPI completion rate is that there are only about 45,000 CAPI cases a month, drastically increasing the variance of any estimates as compared with the mail/CATI cooperation rate (230,000 cases a month). In addition, after calculating the CAPI completion rates, more than 90% of tracts had a completion rate greater than 70%, making it difficult to delineate cut-offs. For these reasons, the CAPI completion rate was not used to redesign the CAPI sampling.

3. ***Will there be a major change to the global design of the CAPI sample?***

Our goal was not to change the basic sample design itself but to assess the effectiveness of the current design and to update the designation of the sampling strata by taking advantage of ACS interview data. The number of CAPI sampling strata as well as the cooperation rate ranges for those strata was not changed. While the design did not change, after the new sampling strata designations were determined, a new reduction factor was calculated to maintain the cost-neutral design.

5.2 Limitations

We were limited in our research by the following:

1. Ideally, weighted observed cooperation rates would be calculated at the tract level using data cumulated over several years of full implementation data collection. However, the differential CAPI sampling rates have only been in use since the selection of the March 2005 CAPI sample corresponding to the January 2005 panel. Therefore the number of sample cases that we were able to include in the analysis was less than optimal.

2. Sample cases in Remote Alaska are not included in this analysis. They were excluded from the original CAPI research also because these cases are unmailable, and therefore contribute zero to the cooperation rate. They are sampled for CAPI at a 2-in-3 rate.

3. Only geocoded records (records for which we know which block it is in) can be included in the cooperation rate calculations. All ungeocoded records are placed into a CAPI sampling stratum, designated as stratum '5'-out of scope, during the initial sample selection and are sampled for CAPI at a rate of 1-in-3 with no reduction. Also, tracts with an unmailable rate greater than 25% are placed into stratum 5 since unmailable addresses are sampled for CAPI at a 2-in-3 rate. Thus, the cooperation rates for CAPI stratum 5 include those tracts where there was either insufficient data available at the time of the initial CAPI strata research.

6. Methodology Used

6.1 Design Area

The design area for the differential CAPI sampling is the Census Tract. The allocation of tracts to the sampling strata produces a file with one record for each of the 65,443 census tracts and one record representing ungeocoded records. This differs from the current methodology in that the initial CAPI research was done using the 2003 current geography and the subsequent allocation of differential CAPI sampling strata was done using current state (based on sample year)/current county (based on sample year)/tract. Due to constantly shifting boundaries, some values of this combination are not on the file used to assign the sampling stratum. By using Census Tract, all addresses will match and will be assigned to a sampling stratum instead of defaulting to CAPI stratum 5 and being sampled at the default rate of 1-in-3.

6.2 Cooperation Rate Calculation

1. Cooperation Rate

We chose a measure of mail/CATI cooperation that does excludes vacants and deletes from the denominator. We should, therefore, be able to better target areas where households are less likely to respond by mail or complete a telephone interview. Response data to calculate the cooperation rates were drawn from the February 2005 through December 2006 monthly panels. January 2005 will not be included due to CAPI workload reductions in some counties while ramping up to full ACS implementation levels.

2. The following components were calculated within each state/county/tract combination:

- Valid, geocoded, and mailable sample addresses, (MAIL): This number is tallied from the January 2005 and January 2006 MAF extracts.

- Valid, geocoded, and mailable vacant sample addresses, (VAC): This number is estimated by applying the vacancy rate to MAIL

- Valid, geocoded, and mailable deleted sample addresses, (DEL): This number is estimated by applying the deletion rate to MAIL

- Weighted number of valid, geocoded, mailable interviews from the mail and CATI modes of data collection, (INTS): This component is calculated by summing the unbiased sampling weights for all mail and CATI interviews.

These components were summed to the Census Tract level within CAPI stratum and the weighted cooperation rate (expressed as a percentage) was calculated:

$$\text{Cooperation Rate} = \left(\frac{\text{INTS}}{\text{MAIL - VAC - DEL}} \right) \times 100$$

6.3 Maintaining Costs

1. Reduction Factor

After allocating tracts to the CAPI strata based on the new cooperation rates, the value of the reduction factor was adjusted for a cost-neutral design. The reduction factor is used to decrease the initial sample size in the highest cooperation rate tracts (CAPI sampling stratum '4') to account for the differential sampling in the low responding tracts. The sample is only reduced in blocks that are densely populated and that are in the two lowest initial sampling strata.

2. Costs

Costs were maintained based upon 2006 response data. The unit costs we used for each mode of data collection are:

MAILC = $12
CATIC = $15
CAPIC = $135

Note that if a sampled address is in the MAIL, CATI, and CAPI phases, then the total cost for that address is cumulative ($162).

3. The following components were calculated at the national level. Group 1 contains all cases where sample reduction does not occur and group 2 contains all cases where sample reduction does occur. Neither group includes Remote Alaska.

- Total number of mailable addresses in 2006 (MAIL)

- Total number of addresses that went to CATI in 2006 (CATI)

- Total number of addresses that went to CAPI in 2006 excluding Remote Alaska (CAPI)

- The estimated number of mail cases in group 'k' (EXM_k)

- The estimated number of CATI cases in group 'k' ($EXCT_k$)

- The estimated number of CAPI cases in group 'k' with the new value of CSTRM ($EXCP_k$)

4. The components listed in 6.3.2 AND 6.3.3 were input into the following cost formula:

$$\text{COST} = \text{MAIL} \times \text{MAILC} + \text{CATI} \times \text{CATIC} + \text{CAPI} \times \text{CAPIC} =$$
$$(\text{EXM}_1 + \text{EXM}_2 \times \text{CFACTOR}) \times \text{MAILC} + (\text{EXCT}_1 + \text{EXCT}_2 \times \text{CFACTOR}) \times \text{CATIC} + (\text{EXCP}_1 + \text{EXCP}_2 \times \text{CFACTOR}) \times \text{CAPIC}$$

This was solved for CFACTOR, which represents the reduction factor and led to the following equation:

$$\text{CFACTOR} = \frac{\text{COST} - \text{EXM}_1 \times \text{MAILC} - \text{EXCT}_1 \times \text{CATIC} - \text{EXCP}_1 \times \text{CAPIC}}{\text{EXM}_2 \times \text{MAILC} + \text{EXCT}_2 \times \text{CATIC} + \text{EXCP}_2 \times \text{CAPIC}}$$

7. Results and Analysis

7.1 Behavior of the Initial CAPI Stratum Designation

We investigated the existing CAPI sampling stratum designation to see if the tract allocation was supported based on the current cooperation rates. We used the same vacancy rate for each tract that was used in the original research. The new CAPI stratum allocation is compared to the old allocation in the following tables.

Table 2. Percentage Distribution of Valid Addresses by Old and New CAPI Sampling Stratum

	New CAPI Sampling Stratum					
	1	2	3	4	5	Total
Old CAPI Sampling Stratum 1	2.11	1.74	0.45	0.38	0.01	4.70
Old CAPI Sampling Stratum 2	1.50	4.23	2.66	2.22	0.01	10.63
Old CAPI Sampling Stratum 3	0.39	2.93	4.07	5.97	0.05	13.42
Old CAPI Sampling Stratum 4	0.41	2.64	6.88	51.90	1.82	63.64
Old CAPI Sampling Stratum 5	0.07	0.14	0.17	0.55	6.70	7.61
Total	4.48	11.68	14.24	61.02	8.57	
Percent No Change						69.01

Table 3. Percentage Distribution of Tracts by Old and New CAPI Sampling Stratum

	New CAPI Sampling Stratum					
	1	2	3	4	5	Total
Old CAPI Sampling Stratum 1	2.96	2.45	0.73	0.59	0.02	6.75
Old CAPI Sampling Stratum 2	1.96	4.63	2.88	2.39	0.02	11.87
Old CAPI Sampling Stratum 3	0.60	3.10	3.83	5.74	0.06	13.33
Old CAPI Sampling Stratum 4	0.73	3.16	7.13	51.51	2.10	64.62
Old CAPI Sampling Stratum 5	0.66	0.13	0.17	0.61	1.87	3.43
Total	6.90	13.46	14.73	60.83	4.07	
Percent No Change						64.79

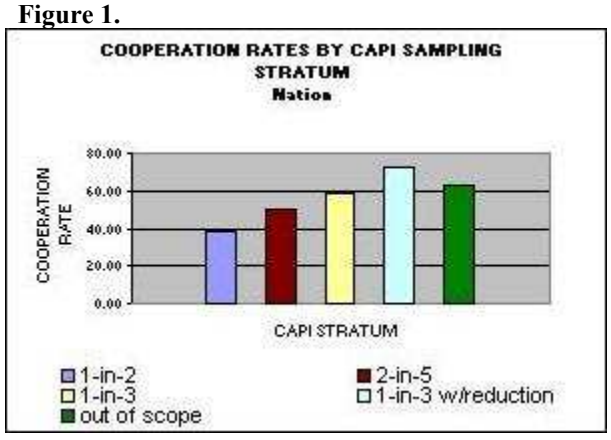
Table 2 shows the distribution of valid addresses from the January 2007 MAF extract by the original allocation and new allocation of the CAPI sampling stratum. More than half of all addresses that were in stratum 1 are now in strata with higher cooperation rates. Both strata 2 and 3 show movement of about 45% of their addresses into higher responding strata, and almost 80% of the addresses in stratum 4 remain in stratum 4. Even though a smaller percentage of addresses are moving out of stratum 4 than the other strata, the larger number of addresses initially in stratum 4 means that the total number of addresses moving isn't that different.

Table 2 also shows the net movement of addresses by showing the total percentage of valid addresses within each stratum for both the current and new allocations. Both strata 1 and 4, the strata with the most extreme rates, lost addresses while the more moderate responding strata, 2 and 3, as well as strata 5, saw a slight increase in their percentages. The last row of Table 2 shows the total percentage of addresses that did not change CAPI stratum. This number is only 69%, meaning that about 31% of addresses would fall into a different CAPI stratum using observed cooperation rates.

Table 3 shows the same distributions as Table 2 for the movement of tracts, not valid addresses. This table shows that a larger percentage of tracts are moving than valid addresses, meaning that the tracts that are moving are smaller tracts with less addresses than the tracts remaining in the same CAPI strata. Stratum 1 even shows a slight increase in total percentage instead of the slight decrease seen in Table 2. This indicates that the valid addresses seen here are in small tracts.

In addition to movement of tracts across CAPI strata, we examined the cooperation rate within current strata at several levels of geography. Figure 1 shows the national level cooperation rate within each stratum. This graph shows the expected distribution with the lowest cooperation in stratum 1 and the cooperation rate increasing for each subsequent stratum. The cooperation rate for stratum 5 is shown for completeness, though tracts in this stratum are out of scope.

Figure 1.



Figures 2 and 3 break the nation into two groups: C2SS counties and expansion counties. These were especially important to look at since only the C2SS counties had data at the time of the original research. The expansion counties had no ACS data; their cooperation rates were modeled from Census 2000 Long Form data. The distribution looks good for the C2SS counties, but the expansion counties show an unexpected distribution. Instead of a steady increase in cooperation rates across all strata, the cooperation rates are relatively flat. This was a concern because it indicates that the current allocation of CAPI strata is over sampling in areas of high response, and that the design could be more efficient.

Figure 2.

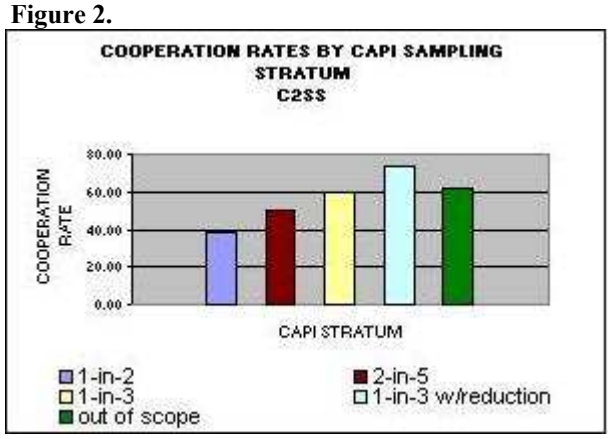
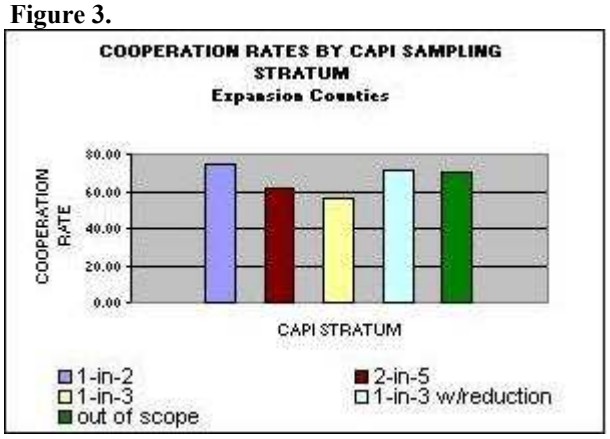


Figure 3.



From Tables 1 and 2, we see that there is quite a bit of change in the distribution of tracts and addresses across the CAPI strata. In addition, in the 1,901 expansion counties, the current CAPI stratum designation is inconsistent with their current cooperation rate. Both of these facts indicate that the allocation of the differential CAPI sampling strata needs to be revised to better reflect the actual ACS cooperation across all tracts.

7.2 Different Research Designs

Since there is now almost two years of ACS data for all counties that use the differential CAPI sampling design, it was possible to use that data to calculate the tract level vacancy rates instead of Census 2000 data. The ACS data is more relevant and more likely to reflect the current state of vacancy across the nation, but since it is survey data it has associated variance.

To see just how much change there would be, the correlation between the two vacancy rates was calculated. With a minimum ACS sample of 1 to calculate a vacancy rate, $\rho = 0.7$. With a minimum sample of 10, $\rho = 0.72$, while with a minimum sample of 50, $\rho = 0.77$. This shows that there is a strong positive correlation, but not so large that no information would be gained from using ACS data.

One way to combat the problem of variance in the ACS vacancy rate estimate is to create a composite vacancy rate by combining it with the Census 2000 vacancy rate. This has the effect of decreasing the variance by 75% while still using the information gained from the ACS. This was done in the following manner:

$$\text{Composite Vacancy Rate} = 0.5 \times (\text{Census VR} + \text{ACS VR})$$

Therefore the variance is:

$$\begin{aligned}
&\text{Var}(\text{Composite VR}) = \text{Var}[0.5 \times (\text{Census VR} + \text{ACS VR})] \\
&= 0.25 \times \text{Var}[(\text{Census VR} + \text{ACS VR})] \\
&= 0.25 \times \text{Var}(\text{Census VR}) + 0.25 \times \text{Var}(\text{ACS VR}) \\
&= 0.25 \times \text{Var}(\text{ACS VR})
\end{aligned}$$

Another factor we considered was whether special treatment outside of the two previous options might be warranted. After reviewing the available information, we decided that using ACS data in the composite vacancy rate would provide a vacancy rate that involved recent information and have an appropriate impact in all areas. In addition, this maintains a consistent approach across the nation.

7.3 Impact on Valid Address and Tract Distributions.

The valid address and tract distributions for the Census Vacancy option can be seen in Tables 2 and 3. The valid address distribution for the Composite Vacancy option can be seen in Table 4. It shows a similar distribution as the Census Vacancy option with almost the same percent of addresses seeing no change. But there is a slight trend toward having fewer addresses in the lower responding strata. Both strata 1 and 2 have about half a percentage point less, and there is slightly more than one percent more addresses in

stratum 4 than the Census Vacancy option. This is because the vacancy rates from ACS tend to be slightly higher than those from Census 2000 resulting in slightly higher cooperation rates.

Table 4. Percentage Distribution of Valid Addresses by Old and New CAPI Sampling Stratum for Composite Vacancy

		New CAPI Sampling Stratum					
		1	2	3	4	5	Total
Old CAPI Sampling Stratum	1	1.98	1.84	0.51	0.37	0.01	4.70
	2	1.33	4.17	2.85	2.27	0.01	10.63
	3	0.31	2.69	4.04	6.33	0.05	13.42
	4	0.34	2.29	6.48	52.72	1.82	63.64
	5	0.07	0.13	0.17	0.55	6.70	7.61
Total		4.03	11.11	14.05	62.24	8.57	
Percent No Change							69.61

Table 5. Percentage Distribution of Tracts by Old and New CAPI Sampling Stratum for Composite Vacancy

		New CAPI Sampling Stratum					
		1	2	3	4	5	Total
Old CAPI Sampling Stratum	1	2.76	2.57	0.80	0.60	0.02	6.75
	2	1.76	4.56	3.09	2.45	0.02	11.87
	3	0.49	2.87	3.85	6.07	0.06	13.33
	4	0.63	2.81	6.73	52.36	2.10	64.62
	5	0.66	0.13	0.17	0.61	1.87	3.43
Total		6.29	12.92	14.63	62.09	4.07	
Percent No Change							65.39

Table 5 shows the tract distributions for the Composite Vacancy option. Overall it shows the same general distribution as the Census Vacancy option in Table 3, but as shown in the valid address distribution, there is a slight change in the total percentage distribution with about 1% more tracts in stratum 4 under the Composite Vacancy option than in the Census Vacancy option.

Even though the overall distributions are close, that doesn't necessarily imply that individual tracts are being allocated to the same CAPI stratum under all designs. For example, Table 6 shows the number of tracts that changed CAPI stratum between the Census Vacancy and Composite Vacancy designs. From the table, 518 tracts moved out of stratum 1, while only 118 moved into stratum 1. So there was a net loss of 400 tracts from stratum 1 by changing from Census Vacancy to the Composite Vacancy option. Similarly, strata 2 and 3 also had a net loss of tracts while stratum 4 gained a total of 823 tracts. This is what we would expect from the previous tables. So while most of the country will not differ between designs, the total number of tracts affected, 4,428, is a significant portion of all tracts at 7%.

Table 6. Tract Count Changes for Each CAPI Sampling Stratum Across Designs

Census Vacancy	Composite Vacancy	Total
1	2	499
1	3	1
1	4	18
2	1	118
2	3	1,146
2	4	8
3	2	407
3	4	1,514
4	2	12
4	3	705
Total		4,428

7.4 Impact on CAPI Workloads

Table 7 shows the national CAPI workload estimates based on 2006 data. It also shows the corresponding reduction factor for each design to maintain costs across the survey. The current reduction factor is 0.92, so there is not much change in any of the designs. This is because even though there is significant change in where specific valid addresses are located, the overall distribution of the nation as a whole did not change a lot for any of the designs.

Table 7. CAPI Workload Estimates and Values of the Reduction Factor

Design	Workload	Reduction Factor
Original	535,963	0.92
Vac Census	537,820	0.91
Vac Composite	536,749	0.92

The Census Vacancy option has the larger increase in the workload of almost 1,900 over the course of the year. That's about 160 cases a month for the whole nation. The Composite Vacancy option had increase in the workload less than half that at about 800 cases a year, or a little more than 65 a month. The differences between the designs will be localized to the 4,458 tracts from the previous table, so while there isn't much difference nationally, there could be an impact at smaller levels of geography.

8. Conclusion

There were some problems evident in the allocation of the initial CAPI stratum designation. In the 1,900 expansion counties, the cooperation rates for the CAPI sampling stratum did not correspond to the expected levels. This was indicative of inefficiencies in the initial CAPI sampling design and is not totally unexpected given the model used.

Both of the designs achieved a similar valid address and tract distribution. The only difference was the slight increase in the percentage of the universe in stratum 4. But even though the overall distributions were similar, there was a moderately sized difference in which tracts were being allocated to which CAPI sampling strata showing that the ACS data used in the Composite Vacancy option did impact the results.

The movement in tracts did exhibit some noticeable changes in the CAPI workloads and the calculation of reduction factor. The Census Vacancy option had a reduction factor around 0.91 while the Composite Vacancy option had a reduction factor around 0.92, which is the current reduction factor. This larger reduction factor also corresponds to a smaller increase in the CAPI workload.

The larger reduction factor for the Composite Vacancy option is also a symptom of having more addresses in stratum 4. This spreads out the sample reduction across more addresses, which reduces its impact on any specific area of the country. The Composite Vacancy reduction factor was the same as the initial reduction factor maintaining consistency across the years.

The Composite Vacancy option takes advantage of the timeliness of the ACS data while lowering the variance with the Census 2000 vacancy rate. The final decision was to use the Composite Vacancy option to reallocate the differential CAPI sampling strata. These CAPI sampling stratum allocations were used for the first time when selecting the 2008 housing unit ACS sample [3].

Acknowledgements

We wish to thank the following people who us helped by assisting in analyzing the data contained in this paper, or helped by providing many clear and useful comments and suggestions: Mark Asiala, Karen E. King, Alfredo Navarro.

References

[1] Asiala, M. (2005). "American Community Survey Research Report: Differential Sub-Sampling in the Computer Assisted Personal Interview Sample Selection in Areas of Low Cooperation Rates". 2005 ACS Documentation Series ACS-DOC-2.

[2] Bates, Lawrence. (2006). "Editing the MAF Extracts and Creating the Unit Frame Universe for the American Community Survey". 2007 ACS Universe Creation Series ACS07-UC-2.

[3] Castro, Edward. (2007). "American Community Survey: Review of the Computer Assisted Personal Interview Sample Design". 2007 ACS Research Series ACS07-R-9.

[4] Hefter, Steven. (2005). "American Community Survey: Specifications for Selecting the Computer Assisted Personal Interview Samples". 2005 ACS Sampling Series ACS-S-45.

[5] Hefter, Steven. (2006). "Specifications for Selecting the Main and Supplemental Housing Unit Address Samples for the American Community Survey". 2007 ACS Sampling Series ACS07-S-3.