Chapter 4.
Sample Design and Selection

4.1   OVERVIEW

The American Community Survey (ACS) and Puerto Rico Community Survey (PRCS) each consist of two separate samples: housing unit (HU) addresses and persons in group quarters (GQ) facilities. As described in Chapter 3, the sampling frames from which these samples are drawn are derived from the Census Bureau's Master Address File (MAF). The MAF is the Census Bureau's official inventory of known living quarters and selected nonresidential units in the United States (U.S.) and Puerto Rico. Independent HU address samples are selected for each of the 3,142 counties and county equivalents in the U.S., including the District of Columbia, for the ACS. Similarly, for the PRCS, address samples are selected for each of the 78 municipalities in Puerto Rico. The first full-implementation county level samples of HU addresses were selected in 2004 and fielded in 2005.[1] Each year, approximately 3 million HU addresses in the U.S. and 36,000 HU addresses in Puerto Rico are selected for the ACS. The first full-implementation samples of GQ facilities and persons were selected independently within each state, including the District of Columbia and Puerto Rico, for use in 2006. In 2006 and 2007, approximately 2.5 percent of the expected number of residents in GQ facilities was included in the ACS and the PRCS, respectively. Beginning in 2008, 16 states with small GQ populations had their sampling rates increased to meet publication thresholds. Details of the data collection methods are provided in Chapters 7 and 8.

This chapter presents details on the selection of the HU address and GQ samples. In some hard-to-reach areas in Alaska, referred to as Remote Alaska, several sampling and data collection processes have been modified. The section on Remote Alaska sampling at the end of this chapter describes the differences in sampling and data collection methodology for Remote Alaska.

4.2   HOUSING UNIT SAMPLE SELECTION

There are two phases of HU address sampling for each county.[2] First-phase sampling includes two stages and involves a series of processes that result in the annual ACS sample of addresses. First-phase sampling is performed twice a year and these two annual processes are referred to as main and supplemental sampling, respectively. During first-phase sampling blocks are assigned to the sampling strata, the sampling rates are calculated, and the sample is selected.[3] During the second phase of sampling, a sample of addresses for which neither a mail questionnaire nor a telephone interview has been completed is selected for Computer Assisted Personal Interviewing (CAPI). This is referred to as the CAPI sample. Figure 4.1 provides a visual overview of the housing unit address sampling process.

First-Phase Sample

The first step of sampling is to assign each address on the sampling frame to one of the five sampling strata by block. Also included in this process are two separate stages of sampling. The first-stage of sampling maintains five distinct partitions of the addresses on the sampling frame for each county. This is accomplished by systematically sorting, and assigning addresses that are new to the frame, to one of the five partitions or subframes[4].
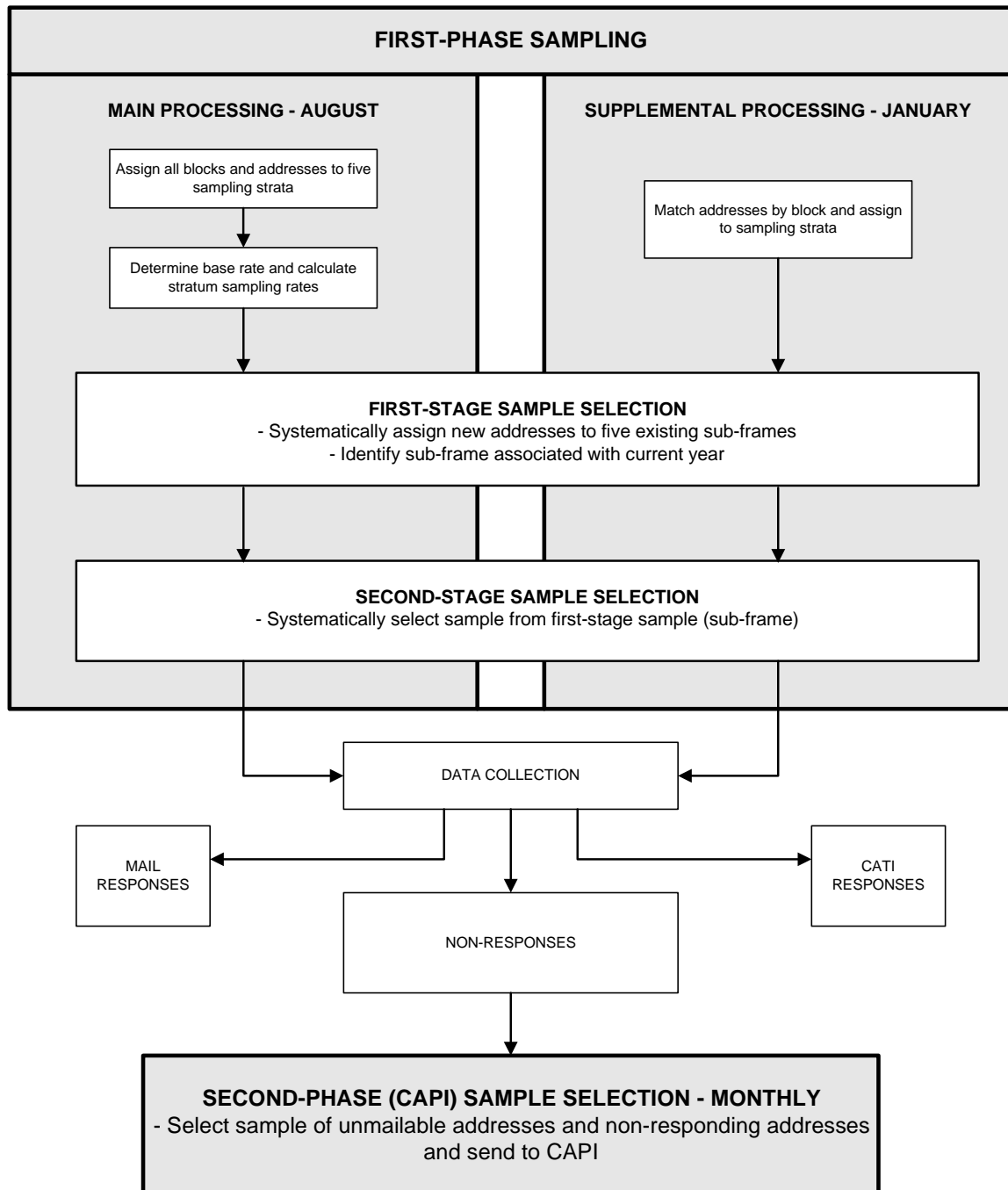
---

[1]In the remainder of this chapter, the term "county" refers to counties, county equivalents, and municipalities.
[2]Throughout this chapter, "addresses" refers to valid ACS addresses that have met the filter criteria (Bates, 2009).
[3]Note that the second-stage sampling rates are calculated once annually during main sampling and these rates are used in supplemental sampling also.
[4]All existing addresses retain their previous assignment to one of the five subframes. The five subframes were created to meet the requirement that no addresses can be in sample more than once in a five-year period.

FIGURE 4.1
SELECTING THE SAMPLES OF HOUSING UNIT ADDRESSES

FIRST-PHASE SAMPLING

MAIN PROCESSING - AUGUST

Assign all blocks and addresses to five sampling strata

Determine base rate and calculate stratum sampling rates

SUPPLEMENTAL PROCESSING - JANUARY

Match addresses by block and assign to sampling strata

FIRST-STAGE SAMPLE SELECTION
- Systematically assign new addresses to five existing sub-frames
- Identify sub-frame associated with current year

SECOND-STAGE SAMPLE SELECTION
- Systematically select sample from first-stage sample (sub-frame)

DATA COLLECTION

MAIL RESPONSES

CATI RESPONSES

NON-RESPONSES

SECOND-PHASE (CAPI) SAMPLE SELECTION - MONTHLY
- Select sample of unmailable addresses and non-responding addresses and send to CAPI

Each subframe is a representative county sample. These subframes have been assigned to specific years and are rotated each year. The subframes maintain their annual designation over time. Finally the sampling rates are determined for each stratum for the current sample year. During the second stage of sampling, a sample of the addresses in the current year's subframe is selected and allocated to the different months for data collection.

Main and Supplemental Sampling

Two separate sampling operations are carried out at different times of the year: (1) main sampling occurs in August and September preceding the sample year, and (2) supplemental sampling occurs in January and February of the sample year. This allows an opportunity for new addresses to have a chance of selection during supplemental sampling. The ACS sampling frames for both main and supplemental sampling are derived from the most recently updated MAF, so the sampling frames for the main and supplemental sample selections differ for a given year. The MAF available at the time of main sampling, obtained in the July preceding the sample year, reflects address updates from October of the preceding year through March of that year. The MAF available at the time of the supplemental sample selection, obtained in January of the sample year, reflects address updates from April through September of the preceding year.

For the main sample, addresses are selected from the subframe assigned to the sample year. These sample addresses are allocated systematically, in a pre-determined sort order, to all 12 months of the sample year. During supplemental sampling, addresses new to the frame are systematically assigned to the five subframes. The new addresses in the current year's subframe are sampled and are systematically assigned to the months of April through December of the sample year for data collection.

Assigning Addresses to the Second-Stage Sampling Strata. Before the first stage of address sampling can proceed for each year's main sampling, each block is assigned to one of the five sampling strata. 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. To accomplish this, each block and its constituent addresses are assigned to one of five sampling strata, each with a unique sampling rate. The stratum assignment for a block is based on information about the set of geographic entities—referred to as sampling entities—which contain the block, or on information about the size of the census tract that the block is located in, as discussed below. Sampling entities are defined as:

- Counties.

- Places with active and functioning governments.[5]

- School districts.

- American Indian Areas/Alaska Native Areas/Hawaiian Home Lands (AIANHH).

- American Indian Tribal Subdivisions with active and functioning governments.

- Minor civil divisions (MCDs) with active and functioning governments in 12 states.[6]

- Census designated places (in Hawaii only).

The sampling stratum for most blocks is based on the measure of size (MOS) for the smallest sampling entity to which any part of the block belongs. To calculate the MOS for a sampling entity, block-level counts of addresses are derived from the main MAF. This count is converted to an estimated number of occupied HUs by multiplying it by the proportion of HUs in the block that were occupied in Census 2000. For American Indian and Alaska Native Statistical Areas (AIANSA[7]) and Tribal Subdivisions, the estimated number of occupied HUs is also multiplied by the proportion of its population that responded as American Indian or Alaska Native (either alone or in combination) in Census 2000. For each sampling entity, the estimate is summed across all blocks

---

[5]Functioning governments have elected officials who can provide services and raise revenue.

[6]The 12 states are considered "strong" MCD states and are: Connecticut, Maine, Massachusetts, Michigan, Minnesota, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont, and Wisconsin.

[7]AIANSA is a general term used to describe American Indian and Alaska Native Village statistical areas. For detailed technical information on the Census Bureau's American Indian and Alaska Native Areas Geographic Program for Census 2000, see the publication in the *Federal Register* (U.S. Census Bureau, 2000).

in the entity and is referred to as the MOS for the entity. In AIANSAs if the sum of these estimates across all blocks is non-zero, then this sum becomes the MOS for the AIANSA. If it is zero (due to a zero census count of American Indians or Alaska Natives), the occupied HU estimate for the AIANSA is the MOS for the AIANSA. For greater detail, see the detailed computer specifications for calculating the MOS for the ACS (Hefter, 2009a). Each block is then assigned the smallest MOS of all the sampling entities in which the block is contained and is referred to as Smallest Entity Measure of Size, or SEMOS.

If the SEMOS is greater than or equal to 1,200, the stratum assignment for the block is based on the MOS for the census tract that contains it. The MOS for each tract (TMOS) is obtained by summing the estimated number of occupied HUs across all of its blocks. Using SEMOS and TMOS, blocks are assigned to the five strata as defined in Table 4.1 below. These strata are consistent with the sampling categories used in Census 2000 except for the category for sampling entities with MOS less than 800 which has been split into two categories for ACS.

Table 4.1 Sampling Strata Thresholds for the ACS/PRCS

Stratum	Smallest Entity Measure of Size (SEMOS) and Tract Measure of Size (TMOS)
Blocks in large sampling entities (SEMOS > 1,200) and large tracts	TMOS > 2,000
Blocks in large sampling entities (SEMOS > 1,200) and small tracts	TMOS ≤ 2,000
Blocks in small sampling entities	800 ≤ SEMOS ≤ 1,200
Blocks in smaller sampling entities	200 ≤ SEMOS < 800
Blocks in smallest sampling entities	SEMOS < 200

Figure 4.2 shows a Census Block that is in City A and also contained in School district 1. Therefore, it is contained wholly in three sampling entities:

- County (not shown)

- Place with active and functioning government—City A

- School district

Example 1: Suppose the MOS for City A is 600 and the MOS for School District 1 is 1,100. Then the SEMOS for the Census Block is 600 and it is placed in the 200 ≤ SEMOS ≤ 800 stratum.

Example 2: Suppose the MOS for City A is 1,300 and the MOS for School District 1 is 1,400, then the SEMOS for the block is 1,300. Since the SEMOS for the block is greater than 1200 the block will be assigned to one of the two strata with SEMOS > 1200 depending on the size of the census tract (TMOS - not shown in the diagram). In this example, suppose the TMOS is 1800, then the Census Block will be placed in the 1200 < SEMOS and TMOS ≤ 2000 stratum.

Determining the Sampling Rates

Each year, the specific set of sampling rates are determined for each of the five sampling strata defined in Table 4.1. Before this can be done, the following three steps are performed. The first step is to calculate a base rate (BR) for the current year. Four of the five sampling rates are a function of a base sampling rate, and the fifth is fixed at ten percent. Table 4.2 shows the relationship between the base rate and the five sampling rates. In 2009, a smaller number of new addresses than was expected were added to the MAF. Therefore a separate set of base rates were calculated for the 2010 supplemental sample selection leading to new supplemental sampling rates.

FIGURE 4.2

ASSIGNMENT OF BLOCKS (AND THEIR ADDRESSES) TO SECOND-STAGE SAMPLING STRATA

(Note that the land area of a sampling entity does not necessarily correlate to its MOS)
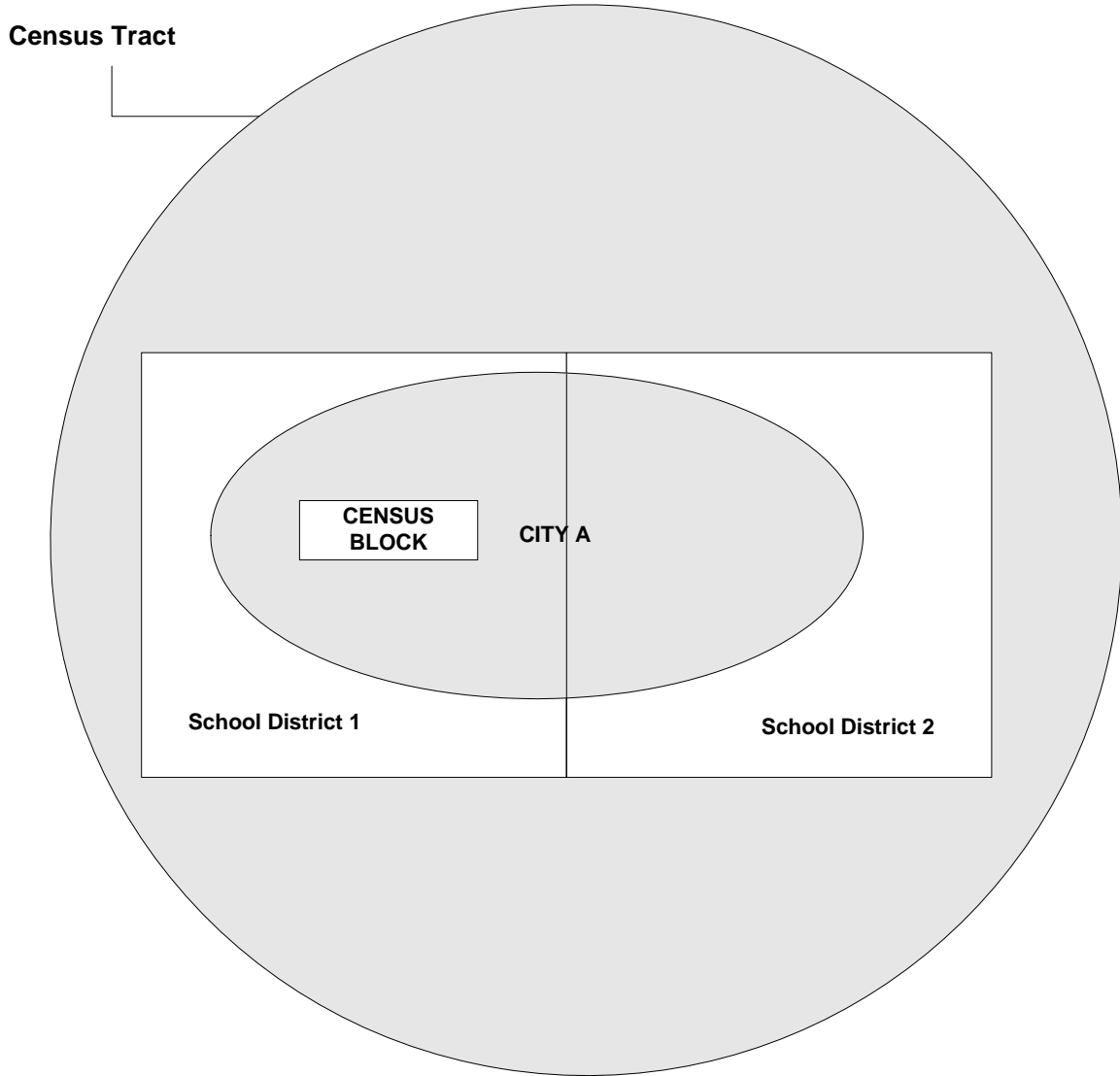
Census Tract

CENSUS BLOCK

CITY A

School District 1

School District 2

Table 4.2 Relationship Between the Base Rate and the Sampling Rates

Stratum	Sampling Rates	
	United States	Puerto Rico
Blocks in large tracts (SEMOS > 1200, TMOS > 2000)............	$0.735 \times BR$	$0.75 \times BR$
Blocks in small tracts (SEMOS > 1200, TMOS ≤ 2000)...........	BR	BR
Blocks in small sampling entities (800 ≤ SEMOS ≤ 1200)......	$1.5 \times BR$	$1.5 \times BR$
Blocks in smaller sampling entities (200 ≤ SEMOS < 800)......	$3 \times BR$	$3 \times BR$
Blocks in smallest sampling entities (SEMOS < 200).............	10 percent	10 percent

The distribution of addresses by sampling stratum, coupled with the target sample size of three million, allows a simple algebraic equation to be set up and solved for BR.

The second step is the calculation of the sampling rates using the value of BR and the equations in Table 4.2. The third step reduces these sampling rates for certain blocks, and is discussed in the following sub-section.

First–Phase Sampling Rates. The sampling rates for the 2009 ACS are given in columns 2 and 4 of Table 4.3 and Table 4.4 for the U.S. and Puerto Rico respectively (Hefter 2009b). Since the design of the ACS calls for a target annual address sample of approximately three million in the U.S. and 36,000 in Puerto Rico, the sampling rates for all but the smallest sampling entities stratum (SEMOS < 200) are reduced each year as the number of addresses in the U.S. and Puerto Rico increases. However, as shown in Table 4.2, among the strata where the rates are decreasing, the relationship of the sampling rates will remain proportionally constant. The sampling rate for the smallest sampling entities will remain at 10 percent.

The sampling rates that are used to select the sample are obtained after the sampling rates are reduced for blocks in specific strata that are in certain census tracts in the U.S. These tracts are predicted to have the highest rates of completed questionnaires by mail and via a telephone follow-up operation, called Computer Assisted Telephone Interviewing (CATI). This adjustment is to compensate for the increase in costs due to increasing the CAPI sampling rates in tracts predicted to have the lowest rate of completed interviews by mail and CATI. Note that the initial identification of these tracts, performed in 2004, was revised in 2007 based on more recent data and was used in the 2008 and 2009 sample selection.

Specifically, the sampling rates are multiplied by 0.92 for some blocks in the U.S. in the two strata in which the SEMOS was greater than 1,200. This adjustment is made for blocks in tracts that were predicted to have a level of completed mail and CATI interviews of at least 60 percent, and at least 75 percent of the block's addresses were defined as mailable.

As a result of this adjustment, there are a total of seven sampling rates used in the U.S., and five in Puerto Rico, as shown in columns 3 and 4 of Table 4.3 and Table 4.4, respectively. See the research report (Asiala, 2005) for a full description of the relationship between this reduction and the CAPI sampling rates. This reduction does not occur in Puerto Rico, so there are five rates used in Puerto Rico.

Table 4.3 2009 ACS/PRCS Main Sampling Rates Before and After Reduction

Stratum (1)	Sampling Rates		
	United States		Puerto Rico
	Before Reduction[1] (2)	After reduction[1] (3)	No Reduction[1] (4)
Blocks in large tracts (SEMOS > 1200, TMOS > 2000)	1.6	(NA)	2.0
Mailable addresses ≥ 75 percent and predicted levels of completed interviews prior to CAPI sampling > 60 percent	(NA)	1.5	(NA)
Mailable addresses < 75 percent or predicted levels of completed interviews prior to CAPI sampling ≤ 60 percent	(NA)	1.6	(NA)
Blocks in small tracts (SEMOS>1200, TMOS≤2000)	2.2	(NA)	2.7
Mailable addresses ≥ 75 percent and predicted levels of completed interviews prior to CAPI sampling > 60 percent	(NA)	2.1	(NA)
Mailable addresses < 75 percent or predicted levels of completed interviews prior to CAPI sampling ≤ 60 percent	(NA)	2.2	(NA)
Blocks in small sampling entities (800 ≤ SEMOS ≤ 1200)	3.3	3.3	4.0
Blocks in smaller sampling entities (200 ≤ SEMOS < 800)	6.5	6.7	8.0
Blocks in smallest sampling entities (SEMOS < 200)	10.0	10.0	10.0

NA Not applicable.

[1] In percent.

Note: The rates in the table have been rounded to one decimal place.

Table 4.4 2009 ACS/PRCS Supplemental Sampling Rates Before and After Reduction

Stratum (1)	Sampling Rates		
	United States		Puerto Rico
	Before Reduction1 (2)	After reduction1 (3)	No Reduction1 (4)
Blocks in large tracts (SEMOS > 1200, TMOS > 2000)	3.0	(NA)	2.0
Mailable addresses ≥ 75 percent and predicted levels of completed interviews prior to CAPI sampling > 60 percent	(NA)	2.8	(NA)
Mailable addresses < 75 percent or predicted levels of completed interviews prior to CAPI sampling ≤ 60 percent	(NA)	3.0	(NA)
Blocks in small tracts (SEMOS>1200, TMOS≤2000)	4.1	(NA)	2.7
Mailable addresses ≥ 75 percent and predicted levels of completed interviews prior to CAPI sampling > 60 percent	(NA)	3.7	(NA)
Mailable addresses < 75 percent or predicted levels of completed interviews prior to CAPI sampling ≤ 60 percent	(NA)	4.1	(NA)
Blocks in small sampling entities (800 ≤ SEMOS ≤ 1200)	6.1	6.1	4.0
Blocks in smaller sampling entities (200 ≤ SEMOS < 800)	12.2	12.2	8.0
Blocks in smallest sampling entities (SEMOS < 200).	18.8	18.8	10.0

NA Not applicable.

[1] In percent.

Note: The rates in the table have been rounded to one decimal place.

First Stage Sample: Random Assignment of Addresses to a Specific Year

One of the ACS design requirements is that no HU address be in a sample more than once in any five-year period. To accommodate this restriction, the addresses in the frame are assigned systematically to five subframes, each containing roughly 20 percent of the frame, and each being a representative sample. Addresses from only one of these subframes are eligible to be in the ACS sample in each year and each subframe is used every fifth year. For example, 2011 will have the same addresses in its subframe as did 2006, with the addition of all new addresses that have been assigned to that subframe during the 2007-2011 time period. As a result, both the main and supplemental sample selection is performed in two stages. The first stage partitions the sampling frame into the five subframes and determines the subframe for the current year, and the second selects addresses to be included in the ACS from the subframe eligible for the sample year.

Prior to the 2005 sample selection, there was a one-time allocation of all addresses then present on the ACS frame to the five subframes. In subsequent years, only addresses new to the frame have been systematically allocated to these five subframes. This is accomplished by sorting the addresses in each county by stratum and geographic order including tract, block, street name, and house number. Addresses are then sequentially assigned to each of the five existing subframes. This procedure is similar to the use of a systematic sample with a sampling interval of five, in which the first address in the interval is assigned to year one, the second address in the interval to year two, and so on. Specifically, during main sampling, only the addresses new to the MAF since the previous year's supplemental MAF are eligible for first-stage sampling and go through the process of being assigned to a subframe. Similarly, during supplemental sampling, only addresses new to the MAF since main sampling go through first-stage sampling. The addresses to be included in the ACS will be selected from the subframe allocated to the sample year during the second-stage of sampling. Additional information can be found in the detailed computer specifications for the HU address sampling (Hefter, 2009c).

Second–Stage Sampling: Selection of Addresses

This sampling process selects a subset of the addresses from the subframe that is assigned to the sample year. This is the final annual ACS sample. These addresses are selected from the subframe in each of the 3,220 counties. The addresses in each county are sorted by stratum and the first-stage order of selection. After sorting, systematic samples of addresses are selected using a sampling rate approximately equal to its final sampling rate divided by 20 percent.[8]

Sample Month Assignment for Address Samples

Each sample address for a particular year is assigned to a data collection month. The set of all addresses assigned to a specific month is referred to as the month's sample or panel. Addresses selected during main sampling are sorted by their order of selection and assigned systematically to the 12 months of the year. However, addresses that have also been selected for one of several Census Bureau household surveys in specified months (which vary by survey) are assigned to an ACS data collection month based on the interview month(s) for these other household surveys.[9] The goal of the assignments is to reduce the respondent burden of completing interviews for both the ACS and another survey during the same month.

The supplemental sample is sorted by order of selection and assigned systematically to the months of April through December. Since this sample is only approximately one percent of the total ACS sample, very few addresses are also in one of the other household surveys in the specified months. Therefore the procedure described above to move the ACS data collection

---

[8]Since the first-stage sampling rate is approximately 20 percent, and the first-stage rate times the second-stage rate equals the sampling rate, the second-stage rate is approximately equal to the sampling rate divided by 20 percent. An adjustment is made to account for uneven distributions of addresses in the subframes.

[9]These surveys include the Survey of Income and Program Participation, the National Crime Victimization Survey, the Consumer Expenditures Quarterly and Diary Surveys, the Current Population Survey, and the State Child Health Insurance Program Surveys.

month for cases in common with the current surveys is not implemented during supplemental first-phase sampling.

4.3   SECOND–PHASE SAMPLING FOR CAPI FOLLOW–UP

As discussed earlier, the ACS uses three modes of data collection—mail, telephone, and personal visit in consecutive months. (See Chapter 7 for more information on data collection.) An interview for a HU and its residents can be completed during the month it was mailed out or during the two subsequent months. All addresses mailed a questionnaire can return a completed questionnaire during this three-month time period.

All mailable addresses with available telephone numbers for which no response is received during the assigned month are sent to CATI for follow-up. The CATI follow-up for these cases is conducted during the following month. Cases where neither a completed mail questionnaire has been received nor a CATI interview completed are eligible for CAPI in the third month, as are the unmailable addresses. An address is considered unmailable if the address is incomplete or directs mail to only a post office box. Table 4.5 summarizes the eligibility of addresses.

Table 4.5 Addresses Eligible for CAPI Sampling

Mailable Address	Responds to Mailing	Responds to CATI	Eligible for CAPI
No	(NA)	(NA)	Yes
Yes	No	No	Yes
Yes	No	Yes	No (completed)
Yes	Yes	(NA)	No (completed)

NA not applicable.

During the CAPI sample selection, a systematic sample of these addresses is selected for CAPI data collection each month, using the rates shown in Table 4.6. The selection is made after sorting within county by CAPI sampling rate, mailable versus unmailable, and geographic order within the address frame (Hefter, 2005).

The variance of estimates for HUs and people living in them in a given area is a function of the number of interviews completed within that area. However, due to sampling for non-response follow-up, CAPI cases generally have larger weights than cases completed by mail or CATI. The variance of the estimates for an area will tend to increase as the proportion of mail and CATI responses decreases. Large differences in these proportions across areas of similar size may result in substantial differences in the reliability of their estimates. To minimize this possibility, tracts in the U.S. that are predicted to have low levels of interviews completed by mail and CATI have their CAPI sampling rates adjusted upward from the default 1-in-3 rate for mailable addresses. This tends to reduce variances for the affected areas both by potentially increasing their total numbers of completed interviews and by decreasing the differences in weights between their CAPI cases and mail/CATI interviews.

No information was available to reliably predict the levels of completed interviews prior to second-phase sampling for CAPI follow-up in Puerto Rico prior to 2005, so the sampling rates of 1-in-3 for mailable and 2-in-3 for unmailable addresses were used initially. On the basis of early response results observed during the first months of the ACS in Puerto Rico, the CAPI sampling rate for mailable addresses in all Puerto Rico tracts was changed to 1-in-2 beginning in June 2005.

Table 4.6 CAPI Sampling Rates

Address and Tract Characteristics	CAPI Sampling rate (percent)
United States	
Unmailable addresses and addresses in Remote Alaska	66.7
Mailable addresses in tracts with predicted levels of completed interviews prior to CAPI subsampling between 0 percent and 35 percent	50.0
Mailable addresses in tracts with predicted levels of completed interviews prior to CAPI subsampling greater than 35 percent and less than 51 percent	40.0
Mailable addresses in other tracts	33.3
Puerto Rico	
Unmailable addresses	66.7
Mailable addresses	50.0

4.4   GROUP QUARTERS SAMPLE SELECTION

GQ facilities include such places as college residence halls, residential treatment centers, skilled nursing facilities, group homes, military barracks, correctional facilities, workers' dormitories, and facilities for people experiencing homelessness. Each GQ facility is classified according to its GQ type. (For more information on GQ facilities, see Chapter 8.) As noted previously, GQ facilities were not included in the 2005 ACS, but have been included each year since 2006. The GQ sample for a given year is selected during a single operation carried out in September and October of the previous year. The sampling frame of GQ facilities and their locations is derived primarily from the most recently available updated MAF and lists from other sources and operations. The ultimate sampling units for the GQ sample are the GQ residents, not the facilities. The GQ samples are independent state level samples. Certain GQ types are excluded from the ACS sampling and data collection operations. These are domestic violence shelters, soup kitchens, regularly scheduled mobile food vans, targeted non-sheltered outdoor locations, commercial maritime vessels, natural disaster shelters, and dangerous encampments. There are several reasons for their exclusion and they vary by GQ type. Concerns about privacy and the operational feasibility of repeated interviewing for a continuing survey, rather than once a decade for a census led to the decision to exclude these GQ types. However, ACS estimates of the total population are controlled to be consistent with the Population Estimates Program estimate of the GQ resident population from all GQs, even those excluded from the ACS.

All GQ facilities are classified into one of three groups: (1) small GQ facilities (having 15 or fewer people according to Census 2000 or updated information); (2) large GQ facilities (with an expected population of more than 15 people); and (3) GQ facilities closed on Census Day (April 1, 2000) or new to the sampling frame since Census Day (with no information regarding the expected population size). There are approximately 105,000 small GQ facilities, 73,000 large GQ facilities, and 2,400 facilities with an unknown population count on the GQ sampling frame. Two sampling strata are created to sample the GQ facilities. The first stratum includes both small GQ facilities and those with no population count. The second includes large facilities. In the remainder of this chapter, these strata will be referred to as the small GQ stratum and the large GQ stratum, respectively. A GQ measure of size (GQMOS) is computed for use in sampling the large GQ facilities. The GQMOS for each large GQ is the expected population count divided by 10.

Different sampling procedures are used for these two strata. GQs in the small GQ stratum are sampled like the HU address sample, and data are collected for all people in the selected GQ facilities. Like HU addresses, small GQ facilities are eligible to be in the sample only once in a five year period. Groups of 10 people are selected for interview from GQ facilities in the large GQ stratum, and the number of these groups selected for a large GQ facility is a function of its

GQMOS. Unlike HU addresses, large GQ facilities are eligible for sampling each year. For greater details, see the computer specifications for the GQ sampling (Hefter, 2009b).

Small Group Quarters Stratum Sample

For the small GQ stratum, a two-phase, two-stage sampling procedure is used. In the first phase, a GQ facility sample is selected using a method similar to that used for the first-phase HU address sample. Just as we saw in the HU address sampling, the first phase has two stages. Stage one systematically assigns small GQ facilities to a subframe associated with a specific year. During the second stage, a systematic sample of the small GQ facilities is selected. In the second phase of sampling, all people in the facility are interviewed as long as there are 15 or fewer at the time of interview. Otherwise, a sub-sample of 10 people is selected and interviewed.

First Phase of Small GQ Sampling—Stage One: Random Assignment of GQ Facilities to Subframes

The sampling procedure for 2006 assigned all of the GQ facilities in the small stratum to one of five 20 percent subframes. The GQ facilities within each state are sorted by small versus closed on Census Day, new versus previously existing, GQ type (such as skilled nursing facility, military barracks, or dormitory), and geographical order (county, tract, block, street name, and GQ identifier) in the small GQ frame. In each year subsequent to 2006, new GQ facilities have been assigned systematically to the five subframes. The subframe for the 2009 GQ sample selection contains the facilities previously designated to the subframe for calendar year 2009 and the 20 percent of new small GQ facilities added since the 2006 sampling activates. The small GQ facilities in the 2009 subframe will not be eligible for sampling again until 2014, since the once-in-five-years period restriction also applies to small GQ facilities.

First Phase of Small GQ Sampling —Stage Two: Selection of Facilities

The second-stage sample is a systematic sample of the GQ facilities from the assigned subframe within each state. The GQs are sorted by new versus previously existing addresses and order of selection. Regardless of their actual size, all of these small GQ facilities have the same probability of selection. The second-stage sampling rate combined with the 1-in-5 first-stage sampling rate yields an overall first-phase-sampling rate equal to the sampling rate for each state. As an example, if the sampling rate for the state is 2.5 percent, then the second-stage sampling rate would be 1-in-8 so that overall the GQ sampling would be (1-in-5) × (1-in-8) = 1-in-40.

Table 4.7 shows the state level sampling rates used beginning in 2008.

Second Stage of Small GQ Sampling: Selection of Persons within Selected Facilities

Every person in the GQ facilities selected in this sample is eligible to be interviewed. If the number of people in the GQ facility exceeds 15, a field sub-sampling operation is performed to reduce the total number of sampled people to 10, similar to the groups of ten selected in the large GQ stratum.

4.5   LARGE GROUP QUARTERS STRATUM SAMPLE

Unlike the HU address and small GQ samples, the large GQ facilities are not divided into five subframes. The ultimate sampling units for large GQ facilities are people, with interviews collected in groups of 10, not the facility itself. A two-phase sampling procedure is used to select these groups: The first indirectly selects the GQ facilities by selecting groups of ten within the facilities and the second selects the people for each facility's group(s) of ten. The number of groups of ten eligible to be sampled from a large GQ facility is equal to its GQMOS. For example, if a facility had 550 people in Census 2000, its GQMOS is 55 and there are 55 groups of ten eligible for selection in the sample.

Table 4.7 2009 Group Quarters Sampling Rates

State	Sampling Rate (percent)	State	Sampling Rate (percent)
Alabama	2.50	Montana	4.38
Alaska	5.53	Nebraska	2.50
Arizona	2.50	Nevada	3.36
Arkansas	2.50	New Hampshire	3.17
California	2.50	New Jersey	2.50
Colorado	2.50	New Mexico	3.06
Connecticut	2.50	New York	2.50
Delaware	4.86	North Carolina	2.50
District of Columbia	3.08	North Dakota	4.59
Florida	2.50	Ohio	2.50
Georgia	2.50	Oklahoma	2.50
Hawaii	3.24	Oregon	2.50
Idaho	3.48	Pennsylvania	2.50
Illinois	2.50	Rhode Island	2.75
Indiana	2.50	South Carolina	2.50
Iowa	2.50	South Dakota	3.91
Kansas	2.50	Tennessee	2.50
Kentucky	2.50	Texas	2.50
Louisiana	2.50	Utah	2.79
Maine	3.14	Vermont	4.95
Maryland	2.50	Virginia	2.50
Massachusetts	2.50	Washington	2.50
Michigan	2.50	West Virginia	2.50
Minnesota	2.50	Wisconsin	2.50
Mississippi	2.50	Wyoming	7.11
Missouri	2.50	Puerto Rico	2.50

First Phase of Large GQ Sampling: Selection of Groups of Ten (and Associated Facilities)

All the large GQ facilities in a state are sorted by GQ type and geographical order in the large GQ frame, and a systematic sample of groups of ten is selected. For this reason, a GQ facility with fewer than 40 groups (or roughly 400 individuals) may or may not have one of its groups selected for the sample. In states with a 2.5 percent sampling rate, GQ facilities with between 40 and 80 groups will have at least one group selected. GQ facilities with between 80 and 120 groups will have at least two groups selected and so forth.

Second Phase of Large GQ Sampling: Selection of Persons within Facilities

The second phase of sampling takes place within each GQ facility that has at least one group selected in the first stage. When a field representative visits a GQ facility to conduct interviews, an automated listing instrument is used to randomly select the 10 people to be included in each group of ten being interviewed. The instrument is pre-loaded with the number of expected person interviews (10 times the number of groups selected), and a random starting number. The field representative then enters the actual number of people in the facility, as well as a roster of their names. To achieve a group size of 10, the instrument computes the appropriate sampling interval based on the observed population at the time of interviewing and then selects the actual people for interviewing using a pre-loaded random start and a systematic algorithm. If the large GQ has

an observed population of 15 or fewer people, the instrument selects a group size of 10 or the observed population if less than 10.

For most GQ types, if multiple groups are selected within a GQ facility, their groups of ten are assigned to different sample months for interviewing. Very large GQ facilities with more than 12 groups selected have multiple groups assigned to some sample months. In these cases, an attempt is made to avoid selecting the same person more than once in a sample month. However, there is no attempt made to avoid selection of someone more than once across sample months within a year. Thus someone in a very large GQ facility could be interviewed in consecutive months. All GQ facilities in this stratum are eligible for selection every year, regardless of their sample status in previous years.

4.6   SAMPLE MONTH ASSIGNMENT FOR SMALL AND LARGE GROUP QUARTER SAMPLES

The selected small GQ facilities and groups of ten for large GQ facilities are assigned to months using a procedure similar to the one used for sampled HU addresses. All GQ samples from a state are combined and sorted by small versus large stratum and first-phase order of selection. Consecutive samples are assigned to the 12 months in a pre-determined order, starting with a randomly determined month.

Due to operational and budgeting constraints, the same month is assigned to all sample groups of ten within certain types of correctional GQs or military barracks. All samples in federal prisons are assigned to September, and data collection may take up to 4.5 months, an exception to the six weeks allowed for all other GQ types. For the samples in non-federal correctional facilities—state prisons, local jails, halfway houses, military disciplinary barracks, and other correctional institutions—or military barracks, individual GQ facilities are randomly assigned to months throughout the year.

4.7   REMOTE ALASKA SAMPLE

Remote Alaska is a set of rural areas in Alaska that are difficult to access and for which all HU addresses are treated as unmailable. Due to the difficulties in field operations during specific months of the year, and the extremely seasonal population in these areas, data collection operations in Remote Alaska differ from the rest of the country. In both the main and supplemental HU address samples, the month assigned for each Remote Alaska HU address is based on the place, AIANSA, block group, or county (in that order) in which it is contained. All designated addresses located in each of these geographical entities are assigned to either January or September. These month assignments are done in such a way as to balance workloads between the months, and to keep groups of cases together geographically. The addresses for each month are sorted by county and geographical order in the address frame, and a sample of 2-in-3 is sent directly to CAPI (no mail or CATI) in the appropriate month. The GQ sample in Remote Alaska is assigned to January or September using the same procedure. Up to four months is allowed to complete the HU and GQ data collection for each of the two data collection periods.

4.8   REFERENCES

Asiala, M. (2005). *American Comunity Survey Research Report: Differential Sub-Sampling in the Computer Assisted Personal Interview Sample Selection in Areas of Low Cooperation Rates.* 2005 American Community Survey Documentation Memorandum Series #ACS05-DOC-2. Washington, DC: U.S. Census Bureau.

Bates, L. (2008). *Editing the MAF Extracts and Creating the Unit Frame Universe for the American Community Survey.* 2009 American Community Survey Universe Creation Memorandum Series #ACS09-UC-1. Washington, DC: U.S. Census Bureau.

Hefter, S. (2005). *American Community Survey: Specifications for Selecting the Computer Assisted Personal Interview Samples.* American Community Survey Sampling Memorandum Series #ACS-S-45. Washington, DC: U.S. Census Bureau.

Hefter, S. (2009a). *Creating the Governmental Unit Measure of Size (GUMOS) Datasets for the American Community Survey and the Puerto Rico Community Survey.* 2009 American Community Survey Sampling Memorandum Series #ACS09-S-1. Washington, DC: U.S. Census Bureau.

Hefter, S. (2009b). *Specifications for Selecting the American Community Survey Group Quarters Sample.* 2009 American Community Survey Sampling Memoradum Series #ACS09-S-6. Washington, DC: U.S. Census Bureau.

Hefter, S. (2009c). *Specifications for Selecting the Main and Supplemental Housing Unit Address Samples for the American Community Survey.* 2009 American Community Survey Sampling Memoradum Series #ACS09-S-3. Washington, DC: U.S. Census Bureau.

U.S. Census Bureau. (2000). American Indian and Alaska Native Areas Geographic Program for Census 2000; Notice. *Federal Register , 65* (121), 39062-39069.