

Chapter 7.

Data Collection and Capture for Housing Units

7.1 OVERVIEW

A key measure of the success of a data collection effort is the final response rate. The American Community Survey (ACS) achieves a high total response rate each year, due in part to the data collection design, which in turn reflects the experience and research in data collection strategies drawn from the U.S. Census Bureau's decennial census and demographic survey programs. Success, however, would not be possible without the high quality of the actual data collection, which is due to the efforts of the interviewing staff in the telephone centers and regional offices. This success also is related to the mandatory nature of the survey. Title 13 of the United States Code [U.S.C.] authorizes the Census Bureau to conduct the ACS, requires households to participate, and requires the Census Bureau to keep confidential all information collected.

The data collection operation for housing units (HUs) consists of three modes: mail, telephone, and personal visit. For most HUs, the first phase includes a questionnaire mailed to the sample address, with a request to the household to complete the questionnaire and return it by mail. If no response is received, the Census Bureau follows up with computer-assisted telephone interviewing (CATI) when a telephone number is available. If the Census Bureau is unable to reach an occupant using CATI, or if the household refuses to participate, the address may be selected for computer-assisted personal interviewing (CAPI).

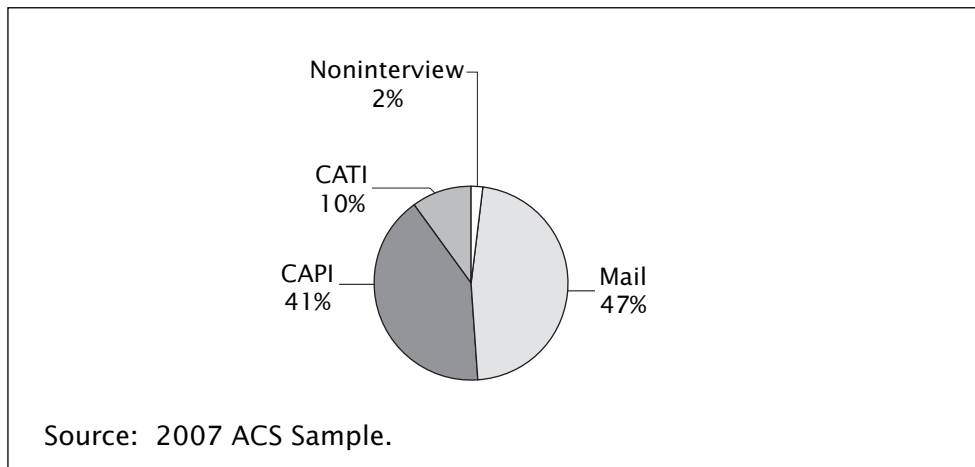
Figure 7.1 ACS Data Collection Consists of Three Overlapping Phases

ACS sample panel	Month of data collection							
	2005		2006					
	November	December	January	February	March	April	May	June
November 2005	Mail	Phone	Personal visit					
December 2005		Mail	Phone	Personal visit				
January 2006			Mail	Phone	Personal visit			
February 2006				Mail	Phone	Personal visit		
March 2006					Mail	Phone	Personal visit	
April 2006						Mail	Phone	Personal visit
May 2006							Mail	Phone
June 2006								Mail

The ACS includes 12 monthly independent samples. Data collection for each sample lasts for 3 months, with mail returns accepted during this entire period, as shown in Figure 7.1. This three-phase process operates in continuously overlapping cycles so that, during any given month, three samples are in the mail phase, one is in the CATI phase, and one is in the CAPI phase.

Figure 7.2 summarizes the distribution of interviews and noninterviews for the 2007 ACS. Among the ACS sample addresses eligible for interviewing in the United States, approximately 47 percent were interviewed by mail, 10 percent by CATI, and 41 percent were represented by CAPI interviews. Two percent were noninterviews.

Figure 7.2 **Distribution of ACS Interviews and Noninterviews**



7.2 MAIL PHASE

Mail is the least expensive method of data collection, and the success of the program depends on high levels of mail response. Sample addresses are reviewed to determine whether the available information is sufficient for mailing. The requirement for a “mailable” address in the United States is met if there is either a complete city-style or rural route address. A complete city-style address includes a house number, street name, and ZIP Code. (The town or city and state fields are not required because they can be derived from the ZIP Code.) A complete rural-route address includes a rural-route number, box number, and ZIP Code. About 95 percent of the 2007 sample addresses in the United States met these criteria and were designated as mailable.

The requirement for a mailable address differs slightly in Puerto Rico. In addition to the criteria for the United States, sample city-style addresses in Puerto Rico also must have an “urbanización” name, building name, or condominium name to be considered mailable. About 72 percent of the addresses in Puerto Rico were considered mailable in 2007.

Examples of unmailable addresses include those with only physical descriptions of an HU and its location, or with post office (P.O.) box addresses, as well as addresses missing place names and ZIP Codes. P.O. box addresses are considered unmailable because of the unknown location of the HU using the P.O. box. Addresses missing ZIP Codes are considered unmailable when the place name is also missing. HU addresses not meeting one of the completeness criteria are still included in the sample frame, but they bypass the mail and telephone phases.

Mailout

Because a high level of mail response is critical, the mail phase used in the ACS consists of three to four mailings to each sample address, depending on when a return is received. ACS materials for U.S. addresses are printed in English, and Puerto Rico Community Survey (PRCS) materials sent to Puerto Rico are printed in Spanish. U.S. respondents can request Spanish mailing packages, and Puerto Rico respondents can request English mailing packages, via telephone questionnaire assistance (TQA). The address label file that includes all mailable sample addresses defines the universe for the first three mailings: a prenotice letter, an initial mail package, and a reminder postcard. A replacement mail package is sent to sample addresses when there is no response 3 weeks after mailing the initial mail package. (Details of each are provided below, and samples are available at <http://www.census.gov/acs/www/SBasics/>.)

Prenotice Letter. The first mailing consists of a prenotice letter, signed by the Census Bureau’s director, alerting residents that they will receive the ACS questionnaire in a few days and encouraging them to return the questionnaire promptly. The prenotice letter is mailed on the Thursday before the last Monday of the month, unless that last Monday is one of the last 2

days of the month, in which case the mailout schedule begins 1 week earlier. The prenotice letter is one of two ACS items printed in-house using print-on-demand technology, which merges the letter text and the sample address from the address label file.

Initial Mail Package. The next mailing is the initial mail package. On the front of the envelope is a boxed message informing recipients that the ACS form is enclosed, and stating in bold, uppercase type that a response is required by law. This initial mail package is mailed on the last Monday of the month or on the previous Monday if the last day of the month is a Monday or a Tuesday. The first mail package includes a cover letter, the questionnaire, an instructional guide, a brochure, and a return envelope.

Cover Letter. The cover letter is signed by the Census Bureau's director. It reminds householders that they received the prenotice letter a few days earlier and encourages them to return the completed questionnaire as soon as possible. The letter then explains the purpose of the ACS and how the data are used. Finally, a toll-free telephone number is included for respondents if they have questions or need help completing the questionnaire.

ACS Questionnaire. The 2006 and 2007 ACS questionnaires are 24-page, two-color booklet-style forms. They are printed on white paper with colored ink—green for the U.S. form, yellow for the Puerto Rico form. The cover of the questionnaire includes information in English and Spanish on how to obtain assistance. The questionnaire includes questions about the HU and the people living in it. Space is provided for detailed information for up to five people. Follow-up by telephone is used for households that return their questionnaires by mail and report that six or more people reside in the household.

Guide to the ACS. The guide instructs respondents how to complete the survey.

Frequently Asked Questions (FAQs) Brochure. This color brochure, available in both English and Spanish, provides answers to frequently asked questions about the ACS. Examples include "What is the American Community Survey?," "Do I have to answer the questions on the American Community Survey?," and "Will the Census Bureau keep my information confidential?" A similar brochure about the PRCS is used in packages mailed to Puerto Rico.

Return Envelope. The postage-paid envelope is for returning the questionnaire to the Census Bureau.

Reminder Postcard. The third mailing is a postcard, also signed by the director of the Census Bureau. The postcard is mailed on Thursdays, 3 days after the initial mail package, and reminds respondents to return their questionnaires. The reminder postcard also is printed in-house, using print-on-demand technology to merge text and addresses.

Replacement Mail Package. The last mailing is sent only to those sample addresses from which the initial questionnaire has not been returned. It is mailed about 3½ weeks after the initial mail package. The contents are the same except that it contains a different cover letter. Signed by the director of the Census Bureau, it reminds the household of the importance of the ACS, and asks them to respond soon.

The Census Bureau's National Processing Center (NPC) assembles and mails the packages for the selected addresses. All of the components of the mail packages except the prenotice letter and reminder postcard are printed under contract by outside vendors. As the vendors print the materials, NPC quality control staff monitor the work and reject materials that do not meet contractual quality standards.

The NPC is responsible for labeling the outgoing mail packages. Several months before each sample's mailings, Census Bureau headquarters staff provides an address file to the NPC for use in creating address labels for the first three mailings. An updated address file is provided to the NPC about 3 days before the mailing of the replacement mail package. This file excludes addresses from which a questionnaire was returned during the first 3 weeks; these usually amount to about 25 to 30 percent of the sample addresses for the United States, and about 10 percent of the sample addresses for Puerto Rico.

Most mail responses are received within 5 weeks after the initial mail package is sent, but the NPC will continue to accept questionnaires for 3 months from the start of each monthly sample. After a specified cutoff date, late mail returns will not be included in the data set.

Check-In

The United States Postal Service (USPS) returns all completed ACS questionnaires to the NPC. The check-in unit receives mail deliveries two or three times each business day. Each questionnaire contains a unique bar code in the address label area. The mail returns are sent through a laser sorter, where the bar code is scanned; this allows sorting by and within monthly sample and by location. During this step, the return envelopes are opened mechanically.

After clerks remove the forms from the return envelopes, the forms are taken to a unit where another set of clerks looks at each page of every returned questionnaire. They also look for enclosed correspondence, which they forward to headquarters, if necessary. The clerks then scan the bar code on each questionnaire to officially check in the form, and organize the forms into batches of 50. Staff have 3 days to check in a form, although usually they check in all the forms they receive within 1 day. Each day, NPC staff transmit a file of the checked-in cases, and headquarters staff update the status of each case in the control file.

Some of the forms are returned to the NPC as “undeliverable as addressed” (UAA) by the USPS. UAAs occur for many reasons, including bad or unknown addresses, vacant HUs, or residents’ refusals to accept mail delivery. Sample addresses that are UAAs initially remain eligible for the replacement mail package because the delivery process for an address often is successful on the second attempt without any change to the address. UAAs are eligible for the CATI and CAPI operations.

Telephone Questionnaire Assistance (TQA)

TQA is a toll-free, interactive voice recognition (IVR) telephone system that respondents can call if they have questions about completing the questionnaire, or to request one in another language. The TQA telephone number is listed on the questionnaire, as well as on all of the letters, brochures, and postcards. Alternate TQA numbers are listed on the questionnaire for Spanish speakers and for a telephone device for the deaf (TDD).

When respondents call TQA, they enter the IVR system, which provides some basic information on the ACS and directions on using the IVR. Respondents may obtain recorded answers to FAQs, or they can speak directly to an agent during business hours. Respondents can furnish their ACS identification number from any of the mailing pieces, which allows them to hear a customized message about the current status of their questionnaire. The IVR can indicate whether the NPC has received a questionnaire for the sample address and, if not, can state that an ACS interviewer may call or visit. If a respondent chooses to speak directly to an agent, the agent answers the caller’s questions and encourages the respondent to complete the questionnaire over the telephone. Agents use an automated survey instrument to capture the respondent’s answers.

Household members from approximately 6 percent of the mailable addresses called the toll-free number for assistance in 2006 and 2007. For less than 1 percent of the mailable addresses in 2006 and 2007, household members agreed to complete the survey over the telephone. All calls are logged, and the system can record up to five reasons for each call. Even though TQA interviews are conducted by telephone, they are considered mail responses because the call was initiated by the sample household upon receiving the questionnaire in the mail.

Data Capture

After the questionnaires have been checked in and batched into groups of 50, they move to the data entry (keying) unit in the NPC. The keying unit has the goal of keying the responses from the questionnaires within 3 weeks of receipt. Data keyers enter the information from the forms into a data capture file. Each day, NPC staff transmit a file with the keyed data, and headquarters staff update the status of each case in the control file. The NPC’s data keying operation uses stringent quality assurance procedures to minimize nonsampling errors.

Data keyers move through three levels of quality assurance verification. When new keyers begin data entry for ACS questionnaires, they are in a training stage, during which 100 percent of their work is checked for correctness. An experienced keyer independently rekeys the same batch of 50 questionnaires, and the work of the two keyers is compared to check for keying errors, defined as incorrectly keyed data items. If the new keyer's error rate (the percentage of all keyed data items that are in error) in one of the first two batches of questionnaires is equal to or less than 1.5 percent, the keyer is moved to the prequalified stage. If the keyer's error rate is greater than 1.5 percent, the keyer is retrained immediately, reassessed, and then advances to the prequalified stage. (These keyers are still subject to 100-percent verification.)

Once prequalified keyers key a batch at an error rate equal to or less than 1.5 percent, they are moved to the qualified stage. If these keyers exceed the error rate of 1.5 percent, they receive immediate feedback. A supervisor eventually decides whether to move them to the qualified stage by verifying a sample of their work, with an acceptable error rate of 1.5 percent or less. Keyers at all levels are subject to removal from the project and administrative action if they fail to maintain an error rate of less than 0.80 percent, but most have a much lower rate.

In mid-2007, the Census Bureau moved to a key-from-image (KFI) data capture system for the HU questionnaires, which involves imaging the questionnaire, interpreting the check box entries with optical mark recognition (OMR), and keying write-in responses from the images using a computerized system. The advantages of KFI include the potential for reduced costs and increased data-capture accuracy.

Failed-Edit Follow-Up

After the data are keyed, the data files are processed in batches through a computerized edit to check coverage consistency and content completeness. This edit identifies cases requiring additional information. Cases that fail are eligible for the telephone failed-edit follow-up (FEFU) operation, and become part of the FEFU workload if a telephone number for the sample address is available. This operation is designed to improve the final quality of mail-returned questionnaires.

Cases failing the edit fall into two broad categories: coverage failures and content failures. Coverage failures can take two forms. First, since the ACS questionnaire is designed to accommodate detailed answers for households with five or fewer people, a case will fail when a respondent indicates that there are more than five people living in the household, or if the reported number of people differs from the number of people for whom responses are provided. Content failures occur if the edit determines that two or more critical items, or a specific number of other required items, have not been answered.

Approximately 33 percent of the keyed mail-return questionnaires in 2006 and 2007 failed either the coverage or content edits and required FEFU. A new set of FEFU cases is generated each business day, and telephone center staff call respondents to obtain the missing data. The interview period for each FEFU case is 3 weeks.

7.3 TELEPHONE PHASE

The second data collection phase is the telephone phase, or CATI. The automated data collection instrument (the set of questions, the list of response categories, and the logic that presents the next appropriate question based on the response to a given question) is written in BLAISE, an open-source scripting software language. The CATI instrument is available in English and Spanish in both the United States and Puerto Rico.

To be eligible for CATI, an HU that did not respond by mail must have a mailable address and a telephone number. The Census Bureau contracts with vendors who attempt to match the ACS sample addresses to their databases of addresses and then provide telephone numbers. There are two vendors for United States addresses and one for Puerto Rico addresses and, since the vendors use different methodologies and sources, one may be able to provide a telephone number while another may not. This matching operation occurs each month before a sample is mailed. About a month later, just prior to the monthly CATI work, headquarters staff transmit a file of the CATI-eligible sample addresses and telephone numbers to a common queue for all three telephone call centers.

The Census Bureau conducts CATI from its three telephone call centers located in Jeffersonville, Indiana; Hagerstown, Maryland; and Tucson, Arizona. The CATI operation begins about 5 weeks after the first mail package is sent out. A control system, WebCATI, is used to assign the cases to individual telephone interviewers. As CATI interviewers begin contacting the households, the WebCATI system evaluates the skills needed for each case (for example, language or refusal conversion skills) and delivers the case to those interviewers who possess the requisite skill(s).

Once a CATI interviewer reaches a person, the first task is to verify that the interviewer has contacted the correct address. If so, the interviewer attempts to complete the interview. If the household refuses to participate in the CATI interview, a different CATI interviewer trained in dealing with refusals will call the household after a few days. If the household again refuses, CATI contact attempts are stopped, and the case is coded as a noninterview. If a household's questionnaire is received at any time during the CATI operation, that case is removed from the CATI sample and is considered a mail response. Each day, NPC staff transmit a file with the status of each case, and headquarters staff update the status on the control file.

The CATI operation has a strong quality assurance program, including CATI software-related quality assurance and monitoring of telephone interviewers. The CATI instrument has a sophisticated, integrated set of checks to prevent common errors. For example, a telephone interviewer cannot input out-of-range responses, skip questions that should have been asked, or ask questions that should have been skipped. Both new and experienced telephone interviewers are subject to random monitoring by supervisors to ensure that they follow procedures for asking questions and effectively probe for answers, and to verify that the answers they enter match the answers provided by the respondent.

Approximately 650 interviewers conduct CATI interviews from the Census Bureau's three telephone call centers. Interviewers participate in a 3-day classroom training session to learn and practice the appropriate interviewing procedures. They have 25 to 26 calendar days to complete the monthly CATI caseload, which averaged in 2006 and 2007 about 95,000 cases each month. At the end of the CATI interview cycle, all cases receive a CATI outcome code in one of three general categories: interview, noninterview, or ineligible for CATI. This last category includes cases with incorrect telephone numbers. Cases in the last two categories are eligible for the personal visit phase.

7.4 PERSONAL VISIT PHASE

The last phase of ACS data collection is the personal visit phase, or CAPI. This phase usually begins on the first day of the third month of data collection for each sample, and typically lasts for the entire month.

After mail and CATI operations have been completed, a CAPI subsample is selected from two categories of cases. Mailable addresses with neither a response to the mailout nor a telephone interview are sampled at a rate of 1 in 2, 2 in 5, or 1 in 3 based on the expected rate of completed interviews at the tract level. Unmailable addresses are sampled at a rate of 2 in 3 (U.S. Census Bureau 2007).

The CAPI operation is conducted by Census Bureau field representatives (FRs) operating from the Census Bureau's 12 regional offices (ROs). The sampled cases are distributed among the 12 ROs based on their geographic boundaries. The Boston RO is responsible for CAPI data collection in Puerto Rico.

After the databases containing the sample addresses are distributed to the appropriate RO, the addresses are assigned to FRs. FRs can conduct interviews by telephone or personal visit, using laptop PCs loaded with a survey instrument similar to the one used in the CATI operation. The CAPI instrument is available in English and Spanish in the United States and Puerto Rico.

If a telephone number is available, the FR will first attempt to call the sample address. There are two exceptions: (1) unmailable addresses, because an FR would not be able to verify the location of the address over the telephone; and (2) refusals from the CATI phase, because these residents already have refused a telephone interview. The FR will call and confirm that he or she has

reached the sample address. If so, the FR uses the automated instrument and attempts to conduct the interview. If an FR cannot reach a resident after calling three to five times at different times of the day during the first few days of the interview period, he or she must make a personal visit.

Approximately 80 percent of CAPI cases require an FR visit. In addition to trying to obtain an interview, a visit is needed to determine whether the HU exists and to determine the occupancy status. If an HU does not exist at the sample address, that status is documented. If an FR verifies that an HU is vacant, he or she will interview a knowledgeable respondent, such as the owner, building manager, real estate agent, or a neighbor, and conduct a “vacant interview” to obtain some basic information about the HU. If the HU is currently occupied, the FR will conduct an “occupied” or “temporarily occupied” interview. An FR conducts a temporarily occupied interview when there are residents living in the HU at the time of the FR’s visit, but no resident has been living there or plans to live there for more than 2 months.

The FRs are trained to remain polite but persistent when attempting to obtain responses. They also are trained in how to handle almost any situation, from responding to a household that claims to have returned its questionnaire by mail to conducting an interview with a non-English speaking respondent.

When FRs cannot obtain interviews, they must indicate the reason. Such noninterviews are taken seriously, because they have an impact on both sampling and nonsampling error. Noninterviews occur when an eligible respondent cannot be located, is unavailable, or is unwilling to provide the survey information. Additional noninterviews occur when FRs are unable to confirm the status of a sample HU due to restricted access to an area because of a natural disaster or nonadmission to a gated community during the interview period. Some sample cases will be determined to be ineligible for the survey. These include sample addresses of structures under construction, demolished structures, and nonexistent addresses.

One of the tasks for an FR is to check the geographic codes (state, county, tract, and block) for each address he or she visits. The FR either confirms that the codes are correct, corrects them, or records the codes if they are missing.

Approximately 3,500 FRs conduct CAPI interviews across the United States and Puerto Rico. Interviewers have almost the entire month to complete the monthly CAPI caseload, which averages more than 40,000 cases each month. Each day, FRs transmit a file with the status of all personal visit cases, and headquarters staff update the statuses on the control file.

FRs participate in a 4-day classroom training session to learn and practice the appropriate interviewing procedures. Supervisors travel with FRs during their first few work assignments to observe and reinforce the procedures learned in training. In addition, a sample of FRs is selected each month and supervisors reinterview a sample of their cases. The primary purpose of the reinterview program is to verify that FRs are conducting interviews, and doing so correctly.

DATA COLLECTION IN REMOTE ALASKA

Remote areas of Alaska provide special difficulties when interviewing, such as climate, travel, and seasonality of the population. To address some of these challenges, the Census Bureau has designated some of these areas to use different procedures for ACS interviewing.

For areas of Alaska that the Census Bureau defines as remote, ACS operations are different from those operations in the rest of the country. The Census Bureau does not mail questionnaires to Remote Alaska sample units and Remote Alaska respondents do not complete any interviews on a paper questionnaire. We do not attempt to conduct interviews with households in Remote Alaska via Census Bureau telephone center interviewers. All interviews for Remote Alaska are conducted using personal visit procedures only.

In order to allow FRs in Alaska adequate time to resolve some of the transportation and logistical challenges associated with conducting interviews in Remote Alaska areas, the normal period for interviewing is extended from 1 month to 4 months. There are two 4-month interview periods every year in Remote Alaska. The first starts in January and stops at the end of April. The second

starts in September and stops at the end of December. These months were identified as most effective in allowing FRs to gain access to remote areas, and in finding residents of Native Villages at home who might be away during the remaining months participating in subsistence activities.

For some boroughs designated as partially remote by the Census Bureau, hub cities in these boroughs are not included in these Remote Alaska procedures. These cities would have cases selected for sample each month of the year, and would be eligible to receive a mail questionnaire, or to be contacted by a telephone center or personal visit interviewer. Table 7.1 provides a list of Remote Alaska areas and their associated interview periods.

Table 7.1 Remote Alaska Areas and Their Interview Periods

Borough name	All or part of borough designated remote	Interview period for the remote portion of the borough	
		January–April	September–December
Aleutians East	All	(X)	
Aleutian Islands	All		(X)
Bethel	Part	½	½
Bristol Bay	All	(X)	
Denali	All		(X)
Dillingham	Part	(X)	
Lake and Peninsula	All		(X)
Nome	Part	½	½
North Slope	Part	(X)	
Northwest Arctic	All	½	½
Southeast	All	½	½
Valdez-Cordova	Part	½	½
Wade Hampton	All	½	½
Yukon-Koyukuk	All	½	½

Note: An X indicates that all workload falls in the interview period.

7.5 REFERENCES

U.S. Census Bureau. 2007. "Accuracy of the Data (2006)." Washington, DC, 2005, <www.census.gov/acs/www/Downloads/ACS/accuracy2006.pdf>.