

ATTACHMENT F
Overview of CPS Sample Design and Methodology

This attachment is an extract of general information about the CPS sample design and weighting from *Current Population Survey, Design and Methodology, Technical Paper 66, October 2006*. Detailed information about these and other aspects of the CPS design is provided in that document, which is available from the Census Bureau Web site:
<http://www.census.gov/prod/2006pubs/tp-66.pdf>.

Chapter 4.

Preparation of the Sample

INTRODUCTION

The Current Population Survey (CPS) sample preparation operations have been developed to fulfill the following goals:

1. Implement the sampling procedures described in Chapter 3.
2. Produce virtually complete coverage of the eligible population.
3. Ensure that only a trivial number of households will appear in the CPS sample more than once over the course of a decade, or in more than one of the household surveys conducted by the U. S. Census Bureau.
4. Provide cost-efficient data collection by producing most of the sampling materials needed for both the CPS and other household surveys in a single, integrated operation.

The CPS is one of many household surveys conducted on a regular basis by the Census Bureau. Insofar as possible, Census Bureau programs have been designed so that survey materials, survey procedures, personnel, and facilities can be used by as many surveys as possible. Sharing personnel and sampling material among a number of programs yields a number of benefits. For example, training costs are reduced when CPS field representatives are employed on non-CPS activities because the sampling materials, listing and coverage instructions, and, to a lesser extent, questionnaire content are similar for a number of different programs. In addition, sharing sampling materials helps ensure that respondents will be in only one sample.

The postsampling codes described in Chapter 3 identify, among other information, the sample cases that are scheduled to be interviewed for the first time in each month of the decade and indicate the types of materials (maps, listing of addresses, etc.) needed by the census field representative to locate the sample addresses. This chapter describes how these materials are put together.

The next section is an overview, while subsequent sections provide a more in-depth description of the CPS sample preparation.

The successful completion of the CPS data collection rests on the combined efforts of headquarters and regional staff. Census Bureau headquarters are located in the Washington, DC, area. Staff at headquarters coordinate

CPS functions ranging from sample design, sample selection, and resolution of subject matter issues to administration of the interviewing staffs maintained under the 12 regional offices, and data processing. Census Bureau staff located in Jeffersonville, IN, also participate in CPS planning and administration. Their responsibilities include preparation and dissemination of interviewing materials, such as maps and segment folders. The regional offices coordinate the interview activities of the interviewing staff.

Monthly sample preparation of the CPS has three major components:

1. Identifying addresses.
2. Listing living quarters.
3. Assigning sample to field representatives.

The within-PSU sample described in Chapter 3 is selected from four distinct sampling frames, not all of which consist of specific addresses. Since the field representatives need to know the exact location of the households or group quarters they are going to interview, much of sample preparation involves the conversion of selected sample information (e.g., maps, lists of building permits) to a set of addresses. This conversion is described below.

Address Identification in the Unit Frame

About 80 percent of the CPS sample is selected from the unit frame. The unit frame sample is selected from a 2000 census file that contains the information necessary for within-PSU sample selection, but does not contain address information. The address information from the 2000 census is stored in a separate file. The addresses of unit segments are obtained by matching the file of 2000 census information to the file containing the associated 2000 census addresses. This is a one-time operation for the entire unit sample and is performed at headquarters. If the addresses are thought to be incomplete (missing a house number or street name), the 2000 census information is reviewed in an attempt to complete the address before sending it to the field representative for interview. In sampling operations, this facilitates the formation of clusters of about 4 housing units, called Ultimate Sampling Units (USUs), as described in Chapter 3.

Address Identification in the Area Frame

About 12 percent of the CPS sample is selected from the area frame. Measures of expected four housing units are selected during within-PSU sampling instead of individual

housing units. This is because many of the addresses in the area frame are not city-style or there is no building permit office coverage. This essentially means that no particular housing units are as yet associated with the selected measure. The only information available is an automated map of the blocks that contain the area segment, the addresses within the block(s) that are in the 2000 census file, the number of measures the block contains, and which measure is associated with the area segment. Before the individual housing units in the area segment can be identified, additional procedures are used to ensure that field representatives can locate the housing units and that all newly built housing units have a probability of selection. A field representative will be sent to canvass the block to create a complete list of the housing units located in the block. This activity is called a listing operation, which is described more thoroughly in the next section. A systematic sampling pattern is applied to this listing to identify the housing units in the area segment that will be designated for each month's sample.

Address Identification in the Group Quarters Frame

About 1 percent of the CPS sample is selected from the group quarters frame. The decennial census files did not have information on the characteristics of the group quarters. The files contain information about the residents as of April 1, 2000, but there is insufficient information about their living arrangements within the group quarters to provide a tangible sampling unit for the CPS. Measures are selected during within-PSU sampling since there is no way to associate the selected sample cases with people to interview at a group quarters. A two-step process is used to identify the group quarters segment. First, the group quarters addresses are obtained by matching to the file of 2000 census addresses, similar to the process for the unit frame. This is a one-time operation done at headquarters. Before the individuals living at the group quarters associated with the group quarters segment can be identified, an interviewer visits the group quarters and creates a complete list of eligible sample units (consisting of people, rooms, or beds) or obtains a count of eligible sample units from a usable register. This is referred to as a listing operation. Then a systematic sampling pattern is applied to the listing to identify the individuals to be interviewed at the group quarters facilities.

Address Identification in the Permit Frame

The proportion of the CPS sample selected from the permit frame increases over the decade as new housing units are constructed. The CPS sample is redesigned about 4 or 5 years after each decennial census, and at this time the permit sample makes up about 6 percent of the CPS sample; this proportion has historically increased about 1 percent a year. Hypothetical measures are selected during within-PSU sampling in anticipation of the construction of new

housing units. Identifying the addresses for these new units involves a listing operation at the building permit office, clustering addresses to form measures, and associating these addresses with the hypothetical measures (or USUs) in the sample.

The Census Bureau conducts the Building Permit Survey, which collects information on a monthly basis from each building permit office (BPO) nationwide about the number of housing units authorized to be built. The Building Permit Survey results are converted to measures of expected four housing units. These measures are continuously accumulated and linked with the frame of hypothetical measures used to select the CPS sample. This matching identifies which BPO contains the measure that is in sample. Using an automated instrument, a field representative visits the BPO to list addresses of units that were authorized to be built; this is the Permit Address Listing (PAL) operation. This list of addresses is transmitted to headquarters, where clusters are formed that correspond one-to-one with the measures. Using this link between addresses and measures, the clusters of four addresses to be interviewed in each permit segment are identified.

Forming clusters. To ensure some geographic clustering of addresses within permit measures and to make PAL listing more efficient, information collected by the Survey of Construction (SOC)¹ is used to identify many of the addresses in the permit frame. The Census Bureau collects information on the characteristics of units to be built for each permit issued by BPOs that are in the SOC. This information is used to form measures in SOC building permit offices. This data is not collected for non-SOC building permit offices.

1. SOC PALs are listings from BPOs that are in the SOC. If a BPO is in the SOC, then the actual permits issued by the BPO and the number of units authorized by each permit (though not the addresses) are known in advance of the match to the skeleton universe. Therefore, the measures for sampling are identified directly from the actual permits. The sample permits can then be identified. These sample permits are the only ones for which addresses are collected.

Because measures for SOC permits were constrained to be within permits, once the listed addresses are complete, the formation of clusters follows easily. The measures formed at the time of sampling are, in effect, the clusters. The sample measures within permits are fixed at the time of sampling; that is, there cannot be any later rearrangement of these units into more geographically compact clusters without voiding the original sampling results.

¹ The Survey of Construction (SOC) is conducted by the Census Bureau in conjunction with the U.S. Department of Housing and Urban Development. It provides current regional statistics on starts and completions of new single-family and multifamily units and sales of new one-family homes.

Even without geographic clustering, there is some degree of compactness inherent in the manner in which measures are formed for sampling. The units drawn from the SOC permits are assigned to measures in the same order in which they were listed in the SOC; thus units within apartment buildings will normally be in the same clusters, and permits listed on adjacent lines on the SOC listing often represent neighboring structures.

2. Non-SOC PALs are listings from BPOS not in the SOC. At the time of sampling, the only data known for non-SOC BPOs is a cumulative count of the units authorized on all permits from a particular office for a given month (or year). Therefore, all addresses for a BPO/date are collected, together with the number of apartments in multiunit buildings. The addresses are clustered using all units on the PAL.

The purpose of clustering is to group units together geographically, thus enabling a reduction in field travel costs. For multiunit addresses, as many whole clusters as possible are created from the units within each address. The remaining units on the PAL are clustered within ZIP Code and permit day of issue.

LISTING ACTIVITIES

When address information from the census is not available or the address information from the census no longer corresponds to the current address situation, then a listing of all eligible units without addresses must be created. Creating this list of basic addresses is referred to as listing. Listing can occur in all four frames: units within multiunit structures, living quarters in blocks, units or residents within group quarters, and addresses for building permits issued.

The living quarters to be listed are usually housing units. In group quarters such as transient hotels, rooming houses, dormitories, trailer camps, etc., where the occupants have special living arrangements, the living quarters listed may be rooms, beds, etc. In this discussion of listing, all of these living quarters are included in the term "unit" when it is used in context of listing or interviewing. Completed listings are sampled at headquarters. Performing the listing and sampling in two separate steps allows each step to be verified and allows more complete control of sampling procedures to avoid bias in designating the units to be interviewed.

In order to ensure accurate and complete coverage of the area and group quarters segments, the initial listing is updated periodically throughout the decade. The updating ensures that changes such as units missed in the initial listing, demolished units, residential/commercial conversions, and new construction are accounted for.

Listing in the Unit Frame

Listing in the unit frame is usually not necessary. The only time it is done is when the field representative discovers

that the address information from the 2000 census is no longer accurate for a multiunit structure and the field representative cannot adequately correct the information.

For multiunit addresses (addresses where the expected number of unit designations is two or more), the field representative receives a segment folder containing a preprinted (computer-generated) Multi-Unit Listing Aid (MULA), Form 11-12, showing unit designations for the segment as recorded in the 2000 census. The MULA displays the unit designations of all units in the structure, even if some of the units are not in sample. Other important information on the MULA includes: the address, the expected number of units at the address, the sample designations, and serial numbers for the selected sample units. If the address is incomplete (missing house number, street name), the field representative receives an Incomplete Address Locator Action Form, which provides additional information to locate the address.

The first time a multiunit address enters the sample, the field representative does one or more of the following:

- For addresses with 2–4 units, verifies that the 2000 census information on the MULA is accurate and corrects the listing sheet when it does not agree with what he/she finds at the address.
- For larger addresses of 5 or more units, resolves missing and duplicate unit designations only for addresses with missing or duplicate unit designations that are in any sample.
- If the changes are so extensive that a MULA cannot handle the corrections, relists the address on a blank Unit/Permit Listing Sheet.

After the field representative has an accurate MULA or listing sheet, he/she conducts an interview for each unit that has a current sample designation. If an address is relisted, the field representative provides information to the regional office on the relisting. The regional office staff will resample the listing sheet (using the sampling pattern on the MULA) and provide the line numbers for the specific lines on the listing sheet that identify the units that should be interviewed.

A regular system of updating the listing in unit segments is not necessary. The field representative may correct an in-sample listing during any visit if a change is noticed.

For single-unit addresses, a preprinted listing sheet is not provided to the field representative since only one unit is expected, based on the 2000 census information. If the address is incomplete (missing house number, no street name), the field representative receives an Incomplete Address Locator Form, which provides additional information to locate the address. If a field representative discovers other units at the address at the time of interview, he/she prepares a Unit/Permit Listing Sheet and lists the

extra units. These additional units, up to and including 15, are interviewed. If the field representative discovers more than 15 units, he/she must contact the regional office for subsampling instructions. For an example of a segment folder, MULA, Unit/Permit Listing Sheet, and Incomplete Address Locator Actions Form, see Appendix A.

Listing in the Area Frame

All blocks that contain area frame sample cases must be listed. Several months before the first area segment in a block is to be interviewed, a field representative visits the block to establish an accurate list of living quarters.

This is a dependent operation conducted via a laptop computer using the Automated Listing and Mapping Instrument (ALMI) software. The field representative starts with a list of the addresses within the block and a digitized map of the block with some addresses mapspotted onto it. The addresses come from the Master Address File (MAF), which is a list of 2000 census addresses, updated periodically through other operations. The digitized map is downloaded from the Census Bureau Topological Integrated Geographic Encoding Reference (TIGER[®]) system.

The field representative updates the block information by matching the living quarters and block features found to the list of addresses and map on the laptop, and makes additions, deletions, or other corrections as necessary. The field representative collects additional data for any group quarters in the block using the Group Quarters Automated Instrument for Listing (GAIL) software.

If the area segment is within a jurisdiction where building permits are issued, housing units constructed since April 1, 2000, are eliminated from the area segment through the year-built procedure to avoid giving an address more than one chance of selection. This is required because housing units constructed since April 1, 2000, Census Day, are already represented by segments in the permit frame. The field representative determines the year each unit was built except for: units in the 2000 census, mobile homes and trailers, group quarters, and nonstructures (buses, boats, tents). To determine "year built," the field representative inquires at each appropriate unit and enters the appropriate information on the computer.

If an area segment is not in a building-permit-issuing jurisdiction, then housing units constructed after the 2000 census do not have a chance of being selected for interview in the permit frame. The field representative does not determine "year built" for units in such blocks.

After the listing of living quarters in the area segment has been completed, the files are transmitted to headquarters where staff then apply the sampling pattern and identify the units to be interviewed.

Periodic updates of the listing are done to reflect change in the housing inventory in the listed block. The following rule is used: The USU being interviewed for the first time

for CPS must be identified from a listing that has been updated within the last 24 months. The field representative updates the area block by verifying the existence of each unit and map feature, accounting for units or features no longer in existence, and adding any new units or features.

Listing in the Group Quarters Frame

The listing procedure is applied to group quarters addresses in the CPS sample that are in the group quarters or area frames. Before the first interviews at a group quarters address can be conducted, a field representative visits the group quarters to establish a list of eligible units (rooms, beds, persons, etc.) at the group quarters. The same procedures apply for group quarters found in the area frame.

Group quarters listing is an independent listing operation conducted via a laptop computer. The instrument, referred to as the Group Quarters Automated Instrument for Listing (GAIL), is used to record the group quarters name, group quarters type, address, the name and telephone number of a contact person, and to list the eligible units within the group quarters, or to obtain a count of the number of eligible units from a register (card file, computer printout, or file) located at the group quarters. Field representatives do not list institutional and military groups quarters; however, they verify that the status has not changed from institutional or military. If it has changed, the field representative will list the addresses of noninstitutional units.

After listing group quarters units or obtaining a count of eligible units from a register, the files are transmitted to headquarters, where staff then apply the sampling pattern to identify the group quarters unit(s) (or units corresponding to sample line(s) within a register) to be interviewed.

The rule for the frequency of updating group quarters listings is the same as for area segments.

Listing in the Permit Frame

There are two phases of listing in the permit frame. The first is the PAL operation, which establishes a list of addresses authorized to be built by a BPO. This is done shortly after the permit has been issued by the BPO and is associated with a sample hypothetical measure. The second listing is required when the field representative visits the unit to conduct an interview.

PAL operation. For each BPO containing a sample measure, a field representative visits the BPO and lists the necessary permit and address information using a laptop computer. If an address given on a permit is missing a house number or street name (or number), then the address is considered incomplete. In this case, a field representative visits the new construction site and draws a Permit Sketch Map showing the location of the structure and, if possible, completes the address.

Permit listing. Listing in the permit frame is necessary for all permit segments. Prior to interviewing, the field representative will receive a segment folder containing a Unit/Permit Listing Sheet (Form 11–3) and, possibly, a Permit Sketch Map to help locate the address. For an example of a segment folder, Unit/Permit Listing Sheets, and a Permit Sketch Map, see Appendix A.

At the time of the interview, the field representative verifies or corrects the basic address. Since the PAL operation does not capture unit designations at a multiunit address, the field representative will list the unit designations prior to interviewing. At both single and multiunit addresses, the field representative will also note any relevant information about the address that would

affect sampling or interviewing, such as conversions, abandoned permits, construction-not-started situations, and more units found than expected for the permit address.

After the field representative has an accurate listing sheet, he/she conducts an interview for each unit that has a current (preprinted) sample designation. If more than 15 units are in sample, the field representative must contact their regional office for subsampling instructions.

The listing of permit segments is not updated systematically; however, the field representative may correct an in-sample listing during any visit if a change is noticed. The change may result in additional units being added or removed from sample.

Chapter 10.

Weighting and Seasonal Adjustment for Labor Force Data

INTRODUCTION

The Current Population Survey (CPS) is a multistage probability sample of housing units in the United States. It produces monthly labor force and related estimates for the total U.S. civilian noninstitutionalized population and provides details by age, sex, race, and Hispanic origin. In addition, estimates for a number of other population sub-domains (e.g., families, veterans, people with earnings, households) are produced on either a monthly or quarterly basis. Each month a sample of eight panels (called rotation groups) is interviewed, with demographic data collected for all occupants of the sample housing units. Labor force data are collected for people 15 years and older. Each rotation group is itself a representative sample of the U.S. population. The labor force estimates are derived through a number of weighting steps in the estimation procedure.¹ In addition, the weighting at each step is replicated in order to derive variances for the labor force estimates. (See Chapter 14 for details.)

The weighting procedures of the CPS supplements are discussed in Chapter 11. Many of the supplements apply to specific demographic subpopulations and differ in coverage from the basic CPS universe. The supplements tend to have higher nonresponse rates.

In order to produce national and state estimates from survey data, a statistical weight for each person in the sample is developed through the following steps, each of which is explained below:

- Preparation of simple unbiased estimates from base-weights and special weights derived from CPS sampling probabilities.
- Adjustment for nonresponse.
- First-stage ratio adjustment to reduce variances due to the sampling of primary sampling units (PSUs).
- National and state coverage adjustments to improve CPS coverage.
- Second-stage ratio adjustment to reduce variances by controlling CPS estimates of the population to independent estimates of the current population.

- Composite estimation using estimates from previous months to reduce the variances.
- Seasonally-adjusted estimates for key labor force statistics.

In addition to estimates of basic labor force characteristics, several other types of estimates are also produced, either on a monthly or a quarterly basis. Each of these involve additional weighting steps to produce the final estimate. The types of characteristics include:

- Household-level estimates and estimates of married couples living in the same household using household and family weights.
- Estimates of earnings, union affiliation, and industry and occupation of second jobs collected from respondents in the quarter sample using the outgoing rotation group's weights.
- Estimates of labor force status by age for veterans and nonveterans using veterans' weights.
- Estimates of monthly gross flows using longitudinal weights.

The additional estimation procedures provide highly accurate estimates for particular subdomains of the civilian noninstitutionalized population. Although the processes described in this chapter have remained essentially unchanged since January 1978, and seasonal adjustment has been part of the estimation process since June 1975, modifications have been made in some of the procedures from time-to-time. For example, in January 1998, a new compositing procedure was introduced; in January 2003, new race cells for the first-stage, second-stage, national, and state coverage steps were added; in January 2005, the number of cells used in the national coverage adjustment and in the second-stage ratio adjustment was expanded to improve the estimates of children.

¹ Weights are needed when the sampled elements are selected by unequal probability sampling. They are also used in poststratification and in making adjustments for nonresponse.