The National Crime Victimization Survey

1The NCVS collects data from residents living throughout the United States, including persons living in non-institutional group quarters, such as dormitories, rooming houses, and religious group dwellings. Crew members of merchant vessels, Armed Forces personnel living in military barracks, and institutionalized persons, such as correctional facility inmates, are not included in the survey. Similarly, U.S. citizens residing abroad and foreign visitors to this country are excluded. With these exceptions, household members age 12 and older, living in units selected for the sample, are eligible to be interviewed.

The NCVS is a panel survey in which the sample is divided into six parts or rotations, each of which contains housing units whose occupants are to be interviewed by telephone or personal visit in a given month and again at 6-month intervals over a period of three years. Each rotation group is further divided into six parts or panels. Persons occupying housing units within a sixth of each rotation group, or one panel, are interviewed each month during the 6-month period. Because the survey is continuous, additional housing units are selected and assigned to rotation groups and panels for subsequent incorporation into the sample. A new rotation group enters the sample every six months, replacing a group phased out after being in the sample for three years.

Sample selection for the NCVS has three stages: the selection of primary sampling units or areas known as PSUs, the selection of address units in sample PSUs, and the determination of persons and households to be included in the sample.

Description of NCVS Sample Frame

Survey estimates are derived from a stratified, multi-stage cluster sample. The PSUs composing the first stage of the sample are formed from counties, groups of adjacent counties, or large metropolitan areas based upon data from the decennial census. The larger PSUs are included in the sample automatically and are considered to be self-representing (SR) since all of them are selected with certainty. The remaining PSUs, called non-self-representing (NSR) because only a subset of them are selected, are combined into strata by grouping PSUs with similar geographic and demographic characteristics, as collected in the 2000 decennial census from which the sample is drawn.

As mentioned previously, the sample design for the NCVS is accomplished in three stages: 1) defining and selecting the PSUs; 2) preparing and selecting address units within the sample PSUs; and 3) determining who at the sample address unit meets the criteria for responding to the survey questions. The NCVS is designed to calculate national estimates of crimes for the target population - the noninstitutional resident population aged 12 years and older. The list used to reach the target population is the

addresses of all living quarters in the U.S. compiled from the most recent decennial census and lists of housing units constructed since that most recent decennial census.

For the 2000-2004 NCVS, sample households were drawn from the 1990-based sample design. For the 2005-2007 NCVS, sample households were drawn from the 1990-based and 2000-based sample designs. For the 2008 and beyond NCVS, sample households were drawn from the 2000-based sample design.

In the second stage of sampling, each selected stratification PSU is divided into four frames (unit, area, permit, and group quarters (GQ)) from which the NCVS sample is independently selected. Small clusters of similar housing units are selected from each frame from each of the selected stratification PSUs. For the unit and GQ frames, addresses come from the 2000 decennial census files. For the permit frame, addresses come from building permit data obtained from building permit offices. For the area frame, sample blocks come from the 2000 decennial census files. Then, addresses are listed and sampled in the field.

For example, approximately 51,000 housing units and other living quarters were designated for the sample in 2008. As explained previously, in order to conduct interviews each month, the sample is divided into six groups, or rotations, and each group of households is interviewed once every 6 months over a period of 3 years. Each rotation group is further divided into six panels. A different panel of households, corresponding to one sixth of each rotation group, is interviewed each month during the 6-month period. Because the survey is continuous, newly constructed housing units are selected as described, and assigned to rotation groups and panels for subsequent incorporation into the sample. A new rotation group enters the sample every 6 months, replacing a group phased out after being in the sample for 3 years.

For these 51,000 sample households, complete interviews were obtained for about 38,064 households in 2008, about 90 percent of all eligible housing units. Within the interviewed households about 78,000 persons, or about 86 percent, provided responses; the other individuals for the most part either refused or were unavailable or unable to answer and no proxy was available or permitted. The remaining 12,845 housing units were not interviewed because they were found to be ineligible -- vacant, demolished, or otherwise ineligible -- for the survey (about 8,816 units), or the occupants could not be reached or refused to participate (about 4,029 units).

1Stage 1. Defining and Selection of PSUs

Defining PSUs
Formation of PSUs begins with listing counties and independent cities in the target area. For the NCVS, this is the entire country.   The counties are either grouped with one or more contiguous counties to form PSUs or are PSUs all by themselves. The groupings are based on certain characteristics such as total land area, current and

projected population counts, large metropolitan areas, and potential natural barriers such as rivers and mountains. The resulting county groupings are called PSUs.

Stratifying PSUs

The NSR PSUs are grouped with similar NSR PSUs within Census divisions (New England, Mid Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, Pacific) to form strata. Each SR PSU forms its own stratum. The data used for grouping the PSUs consist of decennial census demographic data and administrative crime data. As was stated earlier, NSR PSUs are grouped to be as similar or homogeneous as possible. Just as the SR PSUs must be large enough to support a full workload so must each NSR strata be of that size.

Selecting PSUs

The SR PSUs are automatically selected for sample or "selected with certainty." One NSR PSU is selected from each grouped stratum. The NSR PSUs are sampled with probability proportional to the population size using a linear programming algorithm.

Stage 2. Preparing Frames and Sampling Within PSUs

Frame Determination

To ensure adequate coverage for the target population, the Census Bureau defines and selects sample from four address lists called frames: the unit frame, the area frame, the group quarters frame, and the new construction or permit frame. Each address in the country is assigned to one and only one of these frames. Which frame an address is assigned to depends on four factors: 1) what type of living quarters are at the address, 2) when the living quarters were built, 3) where the living quarters were built, and 4) how completely the street address was listed.

Two types of living quarters are defined in the decennial census. The first type is a housing unit. A housing unit (HU) is a group of rooms or a single room occupied as separate living quarters or intended for occupancy as separate living quarters. A housing unit may be occupied by a family or one person, as well as by two or more unrelated persons who share the living quarters. Before the 2000 decennial census, separate living quarters were defined as a space in which the occupants live and eat separately from all the other persons on the property and have direct access to their living quarters from the outside or through a common hall or lobby as found in apartment buildings. Beginning with the 2000 decennial census, the criteria for separate living quarters are that the occupants must live separately from any other individuals in the building and have direct access from outside the building or through a common hall or entry. Eating separately is no longer a criterion.

The second type of living quarters is group quarters (GQ). Group quarters are living quarters where residents share common facilities or receive formally authorized care. About 3 percent of the population counted in the 2000 census resided in group quarters. Of those, less than half resided in non-institutionalized group quarters. About 97 percent of the population counted in the 2000 census lived in housing units. Table 3.1 shows the population distribution by frame in the 2000 decennial census.

Table 1: *Population Distribution by Frame in the 2000 Census*

U.S. Population	Total	281,421,906	100 %
HU Population	Total	273,643,273	97.2 %
GQ Population	Total	7,778,633	2.8%
	Institutional [1]	4,059,039	1.4%
	Non- Institutional GQ Population	3,719,594	1.3%

Unit frame. The unit frame consists of housing units in census blocks that contain a very high proportion of city-style addresses and are essentially covered by building permit offices. The unit frames cover most of the population.

Area frame. The area frame consists of housing units and group quarters in the census blocks that contain a high proportion of non-city-style addresses, or are not covered by building permit offices. Depending on whether or not a block is covered by a building permit office, area frame blocks are classified as area permit or area nonpermit. No distinction is made between area permit and nonpermit blocks during sampling. Group Quarters units can be found in area frame blocks as well as new construction units. In area permit blocks, interviewers attempt to determine what year the house was built to avoid duplication with the Permit Frame.

Group Quarters (GQ) frame. The group quarters frame consists of group quarters in unit frame blocks, and covers a small proportion of the population.

New Construction or Permit frame. The permit frame sampling ensures coverage of housing units built since the most recent decennial census was conducted. The permit frame grows as building permits are issued during the decade. Data collected by the Building Permit Survey are used to update the permit frame monthly.

Within-PSU Sampling
All the Census Bureau's continuing demographic surveys, such as the NCVS, are sampled together shortly after the most recent decennial census. This takes advantage of newly available census data that shows population growth and demographic changes, as well as updated unit address lists. Roughly a decade's worth of sample is selected at that time. Selection of samples is done one survey at a time (sequentially) and one frame at a time (independently). Each survey

---

1[1] Institutional population is not eligible for sample.

determines how the unit addresses within the frame should be sorted prior to sampling. For the NCVS, each frame is sorted by geographic variables. A systematic sampling procedure is used to select living units from each frame. For the unit and the GQ frames, actual unit addresses are selected and reserved for the NCVS. In the area frame, a specified number of living quarters in a specific geographic location are promised to the NCVS and after the address listing operation in that geographic area, the specific unit addresses are assigned. Similarly in the permit frames, empty place holders are selected for the NCVS within the PSU. Then over time as new permits are issued, the place holders are replaced with actual newly built housing units/addresses.

Addresses selected for a survey are removed from the frames, leaving an unbiased or clean universe behind for the next survey that is subsequently sampled. By leaving a clean universe for the next survey, duplication of addresses between surveys is avoided. This is done to help preserve response rates by limiting the number of units falling into more than one survey sample. For the NCVS, roughly 51,000 sample households were selected and contacted twice to collect crime information for collection year 2008. These units were selected based on their geographic and demographic characteristics as determined by the 2000 decennial census.

Stage 3: Sample Within Sample Addresses
The last stage of sampling is done during initial contact of the sample address during the data collection phase. For NCVS, if the address is a residence and the occupants agree to participate, then every person aged 12 or older who lives at the resident address is eligible to be interviewed. For other surveys, only one occupant may be selected. There are procedures to determine who lives in the sample unit and a household roster is completed with their name and other demographic information. For instance, if someone moves out (in) during the interviewing cycle, they are removed from (added to) the roster.

Evaluation of sample frame

Measurement of Error
Undercoverage occurs when some members of the population are inadequately represented in the sample. For example if the NCVS sample for non Black females age 20-21 was 97 percent of the census based estimate for non Black females age 20-21, then the ratio for this subgroup is .97. This is called the coverage ratio. A value less than 1 is an undercoverage ratio and a value greater than 1 is an overcoverage ratio. In general, females tend to have better coverage than males and non Blacks tend to have better coverage than Blacks. The NCVS coverage ratios are comparable to other demographic surveys conducted by the Census Bureau. Average coverage ratios for other Census Bureau demographic surveys are as follows: 71-82 percent for Blacks, 82-90 percent for Non Blacks, 81-88 percent for females, and 83-90 percent for males. This includes such surveys as the Current Population Survey (CPS), the National Health Interview Survey (NHIS), the Survey

of Income and Program Participation (SIPP), and both the Consumer Expenditures Diary (CED) and Quarterly (CEQ) surveys.[2]

Sources of Error in Stage 1 - PSU Selection
If the stratification scheme is not as efficient as promised, it may result in higher sampling errors than expected which may in turn result in the survey not meeting its design goals such as showing that a 5 percent relative consecutive year change was statistically significant at the 95 percent confidence level for certain important estimates. It can also result in differential coverage of subgroups within the target population such as not reaching enough people living in low rent areas or selecting too many people living in Hispanic-headed households.

Sources of Error in Stage 2 - Preparing Frames and Sampling Within PSUs
There are two major sources of bias error in frame determination: over-coverage and under-coverage. Over-coverage exists when a living quarter is listed with more than one unique address. Efforts are made to find and remove duplicate addresses, but some do go undetected. Under-coverage exists when the living quarter is not listed at all. Efforts are made to list all units, but some errors are made. Under-coverage is a bigger problem than over-coverage, so the net result is usually under-coverage. The second stage adjustment in the weighting, which ratio-estimates to independent population controls, was designed to help compensate for either an under-coverage or an over-coverage.

Errors specific to the within PSU sampling phase occur due to incorrect or poor sorting information and incorrect sampling rates. These errors result in too many or too few sample units overall or of a certain type. Incorrect sampling rates are something that can be corrected if detected during sampling verification. However, poor sorting information is less likely to be detected and corrected. For example, suppose a sort is done on the housing units by whether or not the residents owned their home in the last census. Poor or old tenure information about each unit can result in too many sample units owning their home and not enough sample units that rent. This results in an loss of efficiency which increases the variance of estimates.

Sources of Error in Stage 3 - Sample Within Sample Addresses
Being left off the household roster is the most important source of the error in stage 3 sampling for most surveys. Past research states that omissions of persons from rosters in interviewed sample households is the largest source of survey under-coverage. (Shapiro et. al., 1993) Some estimates have it contributing between 60 - 70 percent to overall under-coverage. Since these results were published, work has been done to improve questions about who lives in the sample unit.

_____

[2] 2Source: U.S. Census Bureau (2005) "FY 2003 and 2004 Performance Measures Data for Selected Demographic Surveys." Internal Census Bureau memorandum from Enrique Lamas, September 8,2005.