

# **National Crime Victimization Survey**

Technical Documentation

**NCJ 247252  
September 2014**

## Contents

Chapter 1. NCVS overview .....	1
Purpose and goals.....	1
History.....	2
Key measures .....	3
Personal crime .....	3
Property crime .....	3
NCVS crime taxonomy .....	3
Crime not covered by the NCVS.....	5
Instrument revisions and additions.....	5
Chapter 2. Sample design .....	7
Overview .....	7
Target population.....	7
Sampling frame.....	7
Sample .....	9
Panel design.....	10
First-stage sample design (from 2006 through 2015) .....	12
Coverage.....	13
Living Quarters.....	13
Persons included.....	13
Units/persons excluded by design .....	14
Second-stage sample design (from 2006 through 2015).....	14
Part I. Systematic random sampling from an ordered list .....	15
Part II. Frames for the second-stage sample design .....	16
Contribution of each stage of sampling to the variance .....	20
Sample sizes .....	21
Households .....	21
Persons.....	22
Table 2.5. Distribution of persons in NCVS sample, by sex and race, 2013 .....	22
Sample reinstatement.....	23
Table 2.8. Distribution of NCVS persons from reinstated sample, by sex and race, CY 2013 .....	24
Chapter 3. Data collection.....	25

Field operations .....	25
Field representative training .....	25
Effects of refresher training .....	26
Contact and follow-up procedures .....	26
Survey instruments .....	27
Overview .....	27
NCVS-500 Control Card/household roster and demographic characteristics .....	27
NCVS-1 Basic Screen Questionnaire .....	27
NCVS-2 Crime Incident Report/topical sections .....	28
Supplements.....	28
Interview procedures .....	31
Consent and confidentiality .....	31
Modes .....	32
In-person/telephone .....	32
Use of proxy interviews.....	32
Replacement households .....	35
Spanish and alternative language questionnaires .....	35
Interview auditing and reinterviews .....	36
Nonresponse .....	37
Types of unit nonresponse .....	37
Exclusions due to dangerous areas .....	40
Response rates .....	41
Household response rates .....	41
Person response rates by selected demographic characteristics .....	41
Attrition and time-in-sample (measure of retention over time) .....	42
Response rates by time-in-sample .....	42
Chapter 4. Processing and editing.....	45
Processing overview .....	45
Editing and coding.....	45
Monthly processing, following editing and coding operations .....	46
Quarterly data processing activities.....	47
Imputation.....	48
Chapter 5. Weighting .....	50
Base weights.....	51

Special weighting adjustments .....	51
Adjustment for household nonresponse .....	51
Adjustment for person nonresponse .....	52
Ratio adjustment factors .....	53
First-stage ratio adjustment .....	53
Computing first-stage ratio adjustment factors .....	54
Second-stage ratio adjustment.....	54
Sources of independent controls .....	56
Bounding adjustment.....	56
Adjustment for multiple victims .....	57
Adjustment for series crimes.....	57
Household weight.....	57
Household incident weight.....	57
Design weights for NCVS supplements.....	57
Weighting summaries.....	58
Chapter 6. Variance estimates.....	63
Variance estimates with replication .....	63
Two methods of variance replication .....	63
Replicate weights .....	63
Replicate factors for NSR strata.....	63
Replicate factors for SR strata.....	64
Estimating generalized variance functions.....	67
Generalization method.....	67
The utility of a three-parameter model .....	68
Generalized variances for NCVS supplements .....	68
Year-to-year correlations.....	68
Design effects .....	69
Appendix A: Terms and definitions.....	70
Appendix B: 2013 monthly coverage ratios .....	76
Appendix C: GVF parameters for NCVS .....	88
Appendix D: GVF parameters for NCVS supplements .....	90
Appendix E: Calculating replicate variances with SAS .....	91
NCVS files needed .....	91
Merging files .....	91

SAS code to perform calculations .....	91
Appendix F: 1-year correlations .....	94
Appendix G: 2-year correlations.....	97
Appendix H: Point and variance estimation procedures.....	100
How to estimate statistics with the survey weights.....	103
Rates are not proportions.....	109
How to use Generalized Variance Functions (GVF) to estimate variances .....	109
Why GVFs are used .....	109
How to calculate the variance with a GVF .....	110
Limitations of GVFs.....	111
How to estimate the variance of a difference at two different times.....	111
How to estimate the variance of a rate with a GVF .....	112
How to calculate variance estimates using replicate weights .....	112
References.....	118

## Chapter 1. NCVS overview

### Purpose and goals

The National Crime Victimization Survey (NCVS) is a nationally representative household survey sponsored by the Bureau of Justice Statistics (BJS). It is one of the nation's primary sources of information on criminal victimization, both reported and not reported to police. The NCVS is a self-report survey in which interviewed persons age 12 or older are asked about the number and characteristics of victimizations experienced during the previous 6 months. Households are interviewed every 6 months for a total of seven interviews over 3 years. The survey uses a two-stage approach to identify and enumerate victimizations. In the first stage, respondents are asked a series of screen questions to identify experiences with crime during the 6-month reference period. In the second stage, each victimization identified during the screening process is followed up with a detailed crime incident report that collects information about the date and characteristics of the event.<sup>1</sup> The U.S. Census Bureau serves as the primary data collection organization for the NCVS by conducting interviews and processing sample data on a monthly basis.

Because the NCVS is a self-report survey, only information on nonfatal crimes is collected and murder is excluded. The survey categorizes crimes as “personal” or “property.” Personal crimes (i.e., crimes committed against persons) include rape, sexual assault, robbery, aggravated assault, simple assault, and purse snatching/pocket picking. Property crimes (i.e., crimes committed against a household) include household burglary, motor vehicle theft, and property theft. Data are gathered on types and incidence of crime; monetary losses; physical injuries resulting from crime; characteristics of the victim; and, when appropriate, characteristics of the perpetrator. Periodically, the survey includes separate supplements on additional topics, such as identity theft, crime in schools, workplace violence, and contacts between law enforcement and the public.

The survey was designed with four primary objectives: (1) to develop detailed information about the victims and consequences of crime, (2) to estimate the number and types of crimes not reported to the police, (3) to provide uniform measures of selected types of crimes, and (4) to permit year-to-year comparisons. The survey enables BJS to generate estimates of criminal victimization for the population as a whole and for segments of the population, such as females, the elderly, members of various racial and ethnic groups, location of residence, and other population subgroups.

The NCVS provides the largest national forum for victims to describe their experiences with violent victimization, including information about the offender, such as age, sex, race and ethnicity, and victim–offender relationship; characteristics of the crime event, including time and place, use of weapons, physical injury, and economic consequences of the crime; whether the crime was reported to police; reasons why the crime was reported or not reported to police; and victim experiences with the criminal justice system. Information on all of these elements are collected for each victimization incident.

---

<sup>1</sup>See Appendix A for terms and definitions.

BJS maintains a webpage dedicated to the NCVS on its website: [www.bjs.gov](http://www.bjs.gov).

## History

Between January 1971 and July 1972, the Census Bureau conducted the first nationwide victimization survey as a supplement to the already established Quarterly Household Survey (QHS). During that period, only minor changes were made to the survey questions in an effort to improve data quality.

In July 1972, the National Crime Survey (NCS), as the victimization survey was called before 1991, became a separate national sample survey as a pioneering effort to gather information directly from victims about their experiences with crime. The first sponsor of the NCS was the Law Enforcement Assistance Administration (LEAA), which initiated the survey in response to a mandate set forth by Public Law 93-83 § 515b to collect, evaluate, publish, and disseminate information on the progress of law enforcement within the United States.

The NCS was designed to complement crime data compiled by the FBI and released annually in the *Uniform Crime Reports* (UCR). One of the primary purposes of the UCR is to provide national level estimates of violent and property crime recorded by law enforcement agencies across the United States. However, to appear in the FBI's reporting summary, crimes must have come to the attention of law enforcement *and* been recorded by law enforcement. The NCS was able to provide information on this "dark figure of crime"—as unreported and unrecorded crime came to be known—because the survey was used to interview individuals directly about their experiences with crime. In addition, the NCS provided detailed information about crime that was not available in the UCR, and offered a way to understand the experience of crime from the victim's perspective.

In December 1979, the NCS was transferred to the Bureau of Justice Statistics (BJS) within the U.S. Department of Justice. During that same year, the first steps toward implementing an extensive redesign were undertaken with an eye toward improving the quality and utility of the data collected by the survey.

At the conclusion of the redesign in 1992, the victimization survey was officially renamed the National Crime Victimization Survey (NCVS). Several methodological improvements were implemented to the survey, including—

- an enhanced screening approach using short cues to stimulate respondent recall, thereby improving incident reporting
- screening questions designed to cue respondents on events they may have experienced, rather than relying on subjective interpretations of survey questions
- questions designed to capture additional details on the nature and consequences of victimization that are useful in understanding crime
- questions designed to improve the measurement of sexual and domestic violence.

Additional detail on the redesign may be found in the following documents:

- <http://bjs.ojp.usdoj.gov/content/pub/pdf/ERVE.PDF>
- <http://bjs.ojp.usdoj.gov/content/pub/pdf/ncsrqa.pdf>

In July 2006, the NCVS converted to a fully automated Computer-Assisted Personal Interviewing (CAPI) environment.

## Key measures

### *Personal crime*

Crimes committed against persons are defined as personal crimes and may be violent or nonviolent in nature. Violent crimes involve physical attacks, attempted attacks, and threats of harm, and are always characterized by direct contact between the victim and the offender(s).

Personal crimes that are considered violent, whether attempted or completed, include—

- rape
- sexual assault
- robbery
- aggravated assault
- simple assault
- verbal threats of rape, sexual assault, robbery, and assault.

Rape, sexual assault, robbery, and aggravated assault are considered serious violence. Simple assault is considered a violent crime but does not meet the classification criteria of serious violence.

Nonviolent personal crimes involve personal theft, whether attempted or completed, and are classified under one of the following subcategories:

- purse snatching
- pocket picking.

Personal theft crimes involve an offender taking or attempting to take property or cash directly from the victim by stealth without force or by threat of force. Similar to crimes of violence, personal crimes involve direct contact between an offender(s) and victim during the incident, regardless of whether the crime was completed or attempted. If more than one eligible household member was attacked, verbally threatened, or had property or cash taken directly from them during the same incident, an incident report is completed for each eligible household member who was personally victimized during the incident. (For a representation of person victimizations and incidents, see figure H.1 in Appendix H.)

### *Property crime*

Property crimes include attempted and completed crimes against a household and do not involve direct contact between the offender and a sample household member. Property crimes include—

- household burglary
  - forcible entry burglary
  - unlawful entry without force burglary
- motor vehicle theft
- property theft.

### *NCVS crime taxonomy*

Victimization details collected by NCVS allow crimes to be classified with substantial detail. Table 1 presents the full taxonomy of NCVS crimes.

**Table 1.1. Crime classification taxonomy in the NCVS**

<b>Type of crime</b>
<b>Personal crime</b>
Violent crime
Completed
Attempted
Rape/sexual assault
Rape
Completed
Attempted
Sexual assault
Robbery
Completed
With injury
Without injury
Attempted
With injury
Without injury
Assault
Aggravated
Completed with injury
Attempted/threatened with weapon
Simple
Completed with injury
Attempted/threatened without weapon
Purse snatching/pocket picking
Completed purse snatching
Attempted purse snatching
Pocket picking
<b>Property crime</b>
Burglary
Completed
Forcible entry
Unlawful entry without force
Attempted forcible entry
Motor vehicle theft
Completed
Attempted
Theft
Completed
Less than \$50
\$50-\$249
\$250 or more
Amount not available
Attempted

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2013.

The survey process does not ask respondents which type of crime they experienced, but rather collects the detailed elements of each incident and uses an algorithm to categorize each victimization into a standardized taxonomy. This approach is taken because legal definitions vary by jurisdiction and use of the algorithm allows for a uniform classification of events at the national level.

For example, respondents are not asked if they were robbed. Rather, respondents are asked a series of detailed questions in the NCVS-2 incident report that ascertain whether the victim was present during the incident, whether the victim was directly threatened or attacked, whether the offender used a weapon, and whether the offender took or attempted to take cash or property from the victim. Taken together, these questions provide the criteria required to classify the event as a completed or attempted robbery. Chapter 3 provides additional information about instrumentation and the interview process.

Series crimes can exist for any of the crime types, whether personal or property (assault, robbery, burglary, etc.). Series victimizations are similar in type but occur with such frequency that a victim is unable to recall each individual event or describe each event in detail. Survey procedures allow NCVS interviewers to identify and classify these similar victimizations as series victimizations and to collect detailed information on only the most recent incident in the series.

#### *Crime not covered by the NCVS*

The survey does not measure some types of crimes, including homicide, kidnapping, verbal threats over the phone, and other forms of crime involving social media, arson, fraud, vandalism, drunk driving, and commercial entities. The collection formerly included a survey of commercial entities, but the collection of data on these organizations was dropped in 1977, largely for budgetary reasons. Crimes such as public drunkenness, drug abuse, prostitution, illegal gambling, con games, and blackmail are also not measured.

#### *Instrument revisions and additions*

Content changes have been made to the NCVS questions over the years, which are initiated by external and internal requests.

Questions were added to the incident report section to determine if an incident was a hate crime or was motivated by prejudice or bigotry. In addition, questions were added that ask about any emotional toll the crime incident took and whether a female respondent was pregnant at the time of the incident. A series of questions were added to determine the disability status of the victim, including a question asking if an incident happened because of a respondent's disability. For the 2012 NCVS, the household-level identity theft questions were deleted in favor of collecting such information on a biannual NCVS person-level supplement. To comply with OMB standards, the "race of offender" questions were modified to include more detailed racial and ethnic categories. The table below gives a chronology of questions that were added, revised, and deleted in the NCVS.

**Table 1.2. Chronology of NCVS questions**

January 1999	Hate crime questions added
January 2001	Disability questions added
January 2003	Hate crime questions included on Public Use File (PUF)
July 2004	Household-level identity theft questions added
July 2005	Pregnancy question for female respondents added
January /2008	Disability questions revised for compatibility with American Community Survey (ACS)
July 2008	Vandalism questions deleted
July 2008	Emotional toll questions added
January 2012	Identity theft questions deleted; replaced with ITS supplement
January 2012	Race and ethnicity of offender questions revised

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 1999–2012.

## Chapter 2. Sample design

### Overview

Two types of living quarters are defined in the decennial census. The first type is a housing unit (HU), which 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, one person, or two or more unrelated persons. Before the 2000 decennial census, separate living quarters were defined as a space in which the occupants live and eat separately from all of the other persons on the property and have direct access to their living quarters from outside of the building 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 persons in the building and have direct access to their living quarters from outside of 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), which are living quarters where residents share common facilities or receive formally authorized care. Persons living in GQs are usually not related to each other.

### Target population

The target population of the NCVS is U.S. residents age 12 or older residing in HUs or GQs, such as dormitories, rooming houses, and religious group dwellings. The survey excludes persons under age 12, crew members of maritime vessels; armed forces personnel living in military barracks; the homeless; institutionalized persons, such as correctional facility inmates; U.S. citizens residing abroad; and foreign visitors to the United States. (See page 14 for more information on units/persons excluded by design.) With these exceptions, residents age 12 or older within HUs and GQs selected for the sample are eligible to be interviewed.

The addresses selected for interviewing in the NCVS remain in sample for 3 years, regardless of changes in the HU composition. The composition of some HUs may change through respondent relocation, marriage, divorce, death, and additional changes in household composition, such as the addition of roommates. In addition, younger persons residing in a household will age into the sample when turning age 12. Regardless of these changes in HU composition, all persons living in the HU who are age 12 or older are eligible to be interviewed during each enumeration.

### Sampling frame

To ensure adequate coverage, the Census Bureau defines and selects the NCVS sample from four address lists called frames: (1) unit frame, (2) area frame, (3) group quarters (GQ) frame, and (4) new construction or permit frame. Each address in the United States is assigned to one of these frames, and assignment 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. The main distinction between the frames is the procedure used to collect survey data from that sample address.

**Unit frame.** Consists of HUs in census blocks that contain a very high proportion of city-style addresses and are covered by building permit offices. The unit frame covers most of the population.

**Area frame.** Consists of HUs and GQs in census blocks that either contain a low proportion of 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. GQ units can be found in area frame blocks and new construction units. In area permit blocks, interviewers attempt to determine which year the house was built to avoid duplication with the permit frame. In area frame blocks, addresses are collected and updated through an ongoing listing operation.

**Group quarters (GQ) frame.** Consists of GQs in unit frame blocks. The GQ frame covers a small proportion of the population. Addresses in the GQ frame were extracted from GQs in the 2000 census.

**New construction or permit frame.** Ensures coverage of HUs built since the most recent decennial census. The permit frame grows as building permits are issued during the decade. Data collected by the Census Bureau's Building Permits Survey are used to update the permit frame monthly.

Table 2.1 summarizes the four different frames and the distribution of sampled cases from 2006 through 2013.

**Table 2.1: NCVS Sample distribution of cases by the four frames in 2006-2013**

Criteria for source of the frame	Frames used to cover the U.S.		
	Old construction (HUs built before April 1, 2000)	New construction (HUs built since April 1, 2000)	GQs
<b>Complete addresses with permit office coverage</b>	Unit (76.26%)	Permit (11.22%)	GQ (0.23%)
<b>Otherwise</b>	Area frame (12.29%)		

Note: Detail may not sum to total due to rounding.

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2006–2013.

Whenever possible, the unit/permit/GQ frames are used due to the high cost of the area frame. With the area frame, a field staff member is sent to the blocks that are in the sample and conducts a dependent listing of the units within the block. This means that the field staff starts with computerized maps and address lists that contain all of the known units on the Census Bureau's Master Address File (MAF). The MAF is a continuously updated inventory of all addresses and physical/location descriptions, including their geographic locations, which serves as a source of addresses used for mailing and delivering census forms and for physically locating the addresses when necessary. From there, the field staff adds all of the units they find that are not on the map and removes units on the map that are not on the ground. Maps and addresses are then associated with the units added to the map to select the sample of units for the NCVS interviewer. In

comparison, the unit/permit/GQ frames are less expensive because the unit and GQ frames require no listing.

### *Sample*

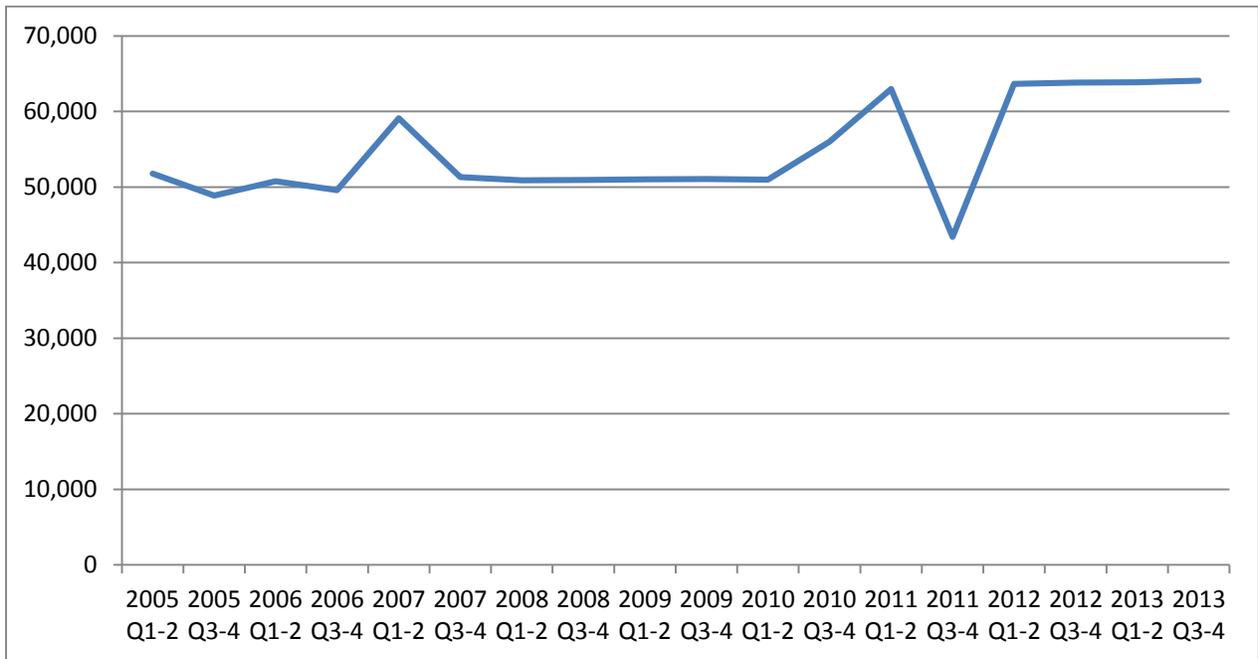
The NCVS uses a two-stage sample design. In the first stage, a sample of primary sample units (PSU) is selected. PSU is a large metropolitan area, county, or group of bordering counties. Within the first stage sample PSUs, a sample of HUs and GQs is selected.

Some PSUs are so large and therefore important to the estimates that they are included in every sample. They are known as self-representing (SR) PSUs because they represent themselves and no other PSUs. Additionally, all PSUs within a large Core-Based Statistical Area (CBSA) are SR PSUs regardless of their size. Other PSUs with smaller populations are stratified within the nine Census divisions: New England, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, and Pacific. These smaller PSUs are grouped with similar PSUs, and one PSU is selected to represent the others included in the stratum. Within each stratum of smaller PSUs, the selection process uses probability proportional to size, where size is based on the number of housing units within the PSU as of the 2000 decennial census. These smaller PSUs that represent themselves and all the other PSUs in the same stratum are referred to as non-self-representing (NSR) PSUs.

All of the sample persons in the survey will come from those geographic areas. Each month, approximately 16<sup>2</sup>/<sub>3</sub>% of sampled HUs or GQs are selected for interviews.

The 2013 NCVS sample contains area, unit, and group quarters frames from the 2000 decennial census. New construction is accounted for in permit frames. The 2013 national sample consisted of approximately 64,000 housing units located throughout the United States during each half-year. Figure 2.1 presents the sample size for each half-year increment from 2005 through 2013. In October 2010, the sample size was increased by reinstating cases from prior sample reductions. In the second half of 2011, data conducted by some field representatives (FR) were excluded from annual estimates at the direction of BJS.

**Figure 2.1. NCVS sample sizes for each half-year, January 2005—December 2013**



Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2005–2013.

**Panel design**

NCVS interviews are conducted continuously throughout the year in a panel design that divides the NCVS sample into six rotating groups. Within each of the six rotating groups, six panels are designated, each of which is interviewed in a given month and every 6 months, either in-person or by telephone, for a total of seven interviews. New sample units rotate into the sample on an ongoing basis to replace outgoing households that have been in sample for the 3-year period. Because the survey is continuous, additional HUs are selected and assigned to rotation groups and panels for subsequent incorporation into the sample. In addition, each rotation group is divided into 6 smaller groups. This approach allows for a new rotation group to enter the sample every 6 months, replacing the group leaving the sample after 3 years, and a new panel within the group enters the sample every month. Figure 2.2 is the NCVS Rotation Chart for January 2012 through December 2014. In the chart, 42 indicates panel 4, rotation group 2.

**Figure 2.2. Panel rotation chart, January 2012–December 2014**

NCVS ROTATION CHART												
January 2012-December 2014												
Year/Month	J25				J26				J27			
2012 JAN	13	14	15	16	11	12	13					
FEB	23	24	25	26	21	22	23					
MAR	33	34	35	36	31	32	33					
APR	43	44	45	46	41	42	43					
MAY	53	54	55	56	51	52	53					
JUNE	63	64	65	66	61	62	63					
JULY		14	15	16	11	12	13	14				
AUG		24	25	26	21	22	23	24				
SEPT		34	35	36	31	32	33	34				
OCT		44	45	46	41	42	43	44				
NOV		54	55	56	51	52	53	54				
DEC		64	65	66	61	62	63	64				
2013 JAN			15	16	11	12	13	14	15			
FEB			25	26	21	22	23	24	25			
MAR			35	36	31	32	33	34	35			
APR			45	46	41	42	43	44	45			
MAY			55	56	51	52	53	54	55			
JUNE			65	66	61	62	63	64	65			
JULY				16	11	12	13	14	15	16		
AUG				26	21	22	23	24	25	26		
SEPT				36	31	32	33	34	35	36		
OCT				46	41	42	43	44	45	46		
NOV				56	51	52	53	54	55	56		
DEC				66	61	62	63	64	65	66		
2014 JAN					11	12	13	14	15	16	11	
FEB					21	22	23	24	25	26	21	
MAR					31	32	33	34	35	36	31	
APR					41	42	43	44	45	46	41	
MAY					51	52	53	54	55	56	51	
JUNE					61	62	63	64	65	66	61	
JULY						12	13	14	15	16	11	12
AUG						22	23	24	25	26	21	22
SEPT						32	33	34	35	36	31	32
OCT						42	43	44	45	46	41	42
NOV						52	53	54	55	56	51	52
DEC						62	63	64	65	66	61	62

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2012–2014.

Before 2007, the initial interview was excluded in the estimates. This interview was unbounded, meaning it did not have a prior interview to frame the reference period. In order to minimize over-reporting, the first interview was not used in estimates and was only used as a bounding interview for the second interview. Bounding is a process to ensure that previously reported incidents are not reported again in the enumeration that follows. This is known as telescoping, or bringing incidents outside the reference period into the period. Bounding provides a more accurate measure of criminal victimization within the NCVS sample HU or GQ. For budgetary reasons, since January 2007, the bounding interview has been included in the computation of annual estimates. A weighting adjustment is applied to compensate for the use of the first interview and the potential for victimizations that are out of the reference period.

NCVS is a panel survey in which each HU or GQ is interviewed every 6 months, for a total of seven interviews. NCVS has a panel design because panel surveys have small variances for estimates of year-to-year change, compared to a sample design that selects two independent samples in separate years. A longitudinal design also enables researchers to examine within person and household change over time, though BJS typically does not analyze the data in this manner. In addition, a panel design was chosen for the advantages of cost and practicality. A

panel survey allows for the measurement of change in a unit between two periods of interest (i.e., year-to-year change) because most of the units are interviewed during both periods. Unlike two independent samples, the observations from the same unit are correlated, and this correlation makes the variances smaller.

For NCVS, about 60% of the sample units will be in common for an estimate that involves the annual estimates of 2 consecutive years. Figure 2.2 provides an illustration of this overlap. For example, the estimator of the difference between two annual estimates will have 60% of the sample common to both annual estimates. This amount of overlap assumes that the sample is in a “steady state”—no reductions or expansions occurred during either year, nor was either year part of a phase-in/phase-out period between sample designs, or any other unusual circumstances other than the steady state sample design. Similarly, estimators involving two annual estimates that are 2 years apart have about 33% of the sample units contributing to both of the annual estimates of the difference.

When the residents of a sampled address move away and are replaced by new residents, the new household is considered a replacement household. A household roster is administered to the new residents and interviewing continues in the usual manner with eligible members of the replacement household. Historically, replacement households entering the sample during the second through seventh interview have not been bounded, and current procedures do not adjust victimization estimates generated from replacement households.

### **First-stage sample design (from 2006 through 2015)**

This section describes the basic components of the first-stage sample design for NCVS. The information is specific to the 2000 sample design, which includes estimates from 2006 through 2015. However, much of the 2000 sample design is similar to the 1990 sample design, which is associated with NCVS estimates from 1996 through 2005.

The first stage of sample design involves dividing the United States into 2,025 first-stage sample units or PSUs consisting of large metropolitan areas, counties, or groups of bordering counties. These areas are grouped together using data from the Census 2000 on characteristics such as total land area, current and projected population counts, large metropolitan areas, and natural barriers, such as rivers and mountains. The desirable population size for a PSU was at least 7,500 persons. The intended maximum land area was kept at 3,000 square miles. However about one in five PSUs exceeded this threshold because many counties had a large geographic area but a small population. These limits were chosen to provide sufficient workload for at least one field representative per PSU while ensuring travel time for individual FRs was not overly onerous.

As stated earlier, the NSR PSUs are stratified within the nine census divisions and grouped with similar NSR PSUs to form strata.

Eight variables are used to create the strata, including—

- average number of reported crimes (except aggravated assault) for the years from 1986 through 1996 (Uniform Crime Reporting data)
- HUs with a Hispanic or Latino head of household
- population in same house as previous 5 years
- occupied HUs with an income less than \$15,000
- renter-occupied HUs in 5 or more unit structure
- urban HUs

- owner-occupied HUs with a value less than or equal to \$70,000
- renter-occupied HUs with rent less than \$450.

PSUs may vary on a number of characteristics, most importantly by size. For this reason, the sample of PSUs with unequal probabilities of selection is chosen. As with most household surveys, NCVS selects PSUs with probabilities proportional to a measure of size (MOS), where the MOS for NCVS is the population in the PSU.

Some of the PSUs selected in the 1990 sample design were not selected in the 2000 sample design and vice versa. PSUs selected in two consecutive first-stage sample designs are referred to as PSU overlap. Mathematical programming techniques in use since the 1970 redesign maximize the probability of selecting PSUs that are already in sample while maintaining the correct overall probability of selection. This approach maximizes overlap, which minimizes the number of new PSUs during any given sample redesign. It also minimizes the hiring and training of new field representatives and concomitant loss of accuracy resulting from the hiring and training of new field representatives. Though PSU overlap is maximized, the correct probability of selection is maintained and probabilities of selection remain proportional to the MOS.

## Coverage

### *Living Quarters*

Two types of living quarters (HU and GQ) are defined in the decennial census. (See the Overview in this chapter or Appendix A for further information on these types of living quarters.) About 3% of the population counted in the 2010 census resided in GQs. Of those, about half resided in non-institutionalized GQs.

### *Persons included*

Three types of respondents are included in the NCVS:

- household respondents
- individual respondents,
- proxy respondents.

Certain questions within the NCVS instrument are designed for different respondents. Questions relating to crimes affecting the HU are asked only once during each enumeration period. The respondent for these questions is referred to as the household respondent.

Other questions in the NCVS are considered self-response questions and relate to crimes affecting persons and not the HU. In other words, each HU member age 12 or older is expected to answer for himself or herself. These respondents are referred to as individual respondents.

As a last resort and only under specific conditions, another person is allowed to answer questions for a HU member. This person is referred to as a proxy respondent. In most cases, a proxy respondent will be another HU member. Strict rules are in place for when to accept a proxy interview because a proxy respondent is less likely to report a crime incident and be knowledgeable of the details concerning a victimization incident. Table 2.2 provides a summary of the total number of households and persons in the half-year samples for 2013, the total number of households and persons that were interviewed, the total number of household and person incidents reported (incidents that occurred within the United States), and the total number of proxy respondents in calendar year 2013.

**Table 2.2. Incidents reported in the NCVS half-year sample, by total households and persons, interviews, and proxy respondents, 2013**

	Quarter 1–2, 2013			Quarter 3–4, 2013			2013
	Total	Interviews	Incidents reported	Total	Interviews	Incidents reported	Proxy respondents
Households	63,871	45,759	3,172	64,078	44,870	3,017	N/A*
Persons	92,306	80,545	781	90,393	79,499	695	7,675

\*Household questions are not administered to proxy respondents.

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2013.

### *Units/persons excluded by design*

The survey excludes HU members under age 12; armed forces personnel living in military barracks; the homeless; institutionalized persons, such as correctional facility inmates; U.S. citizens residing abroad; and foreign visitors to the United States.

Additionally, persons living in GQs built after April 1, 2000, in unit frame blocks are excluded. The following types of GQs are also excluded, regardless of whether persons reside in these locations:

- military ships
- shelters for abused women
- mental hospitals
- soup kitchens
- regularly scheduled mobile food vans
- targeted nonsheltered outdoor locations
- crew members of maritime vessels
- GQs for victims of natural disasters.

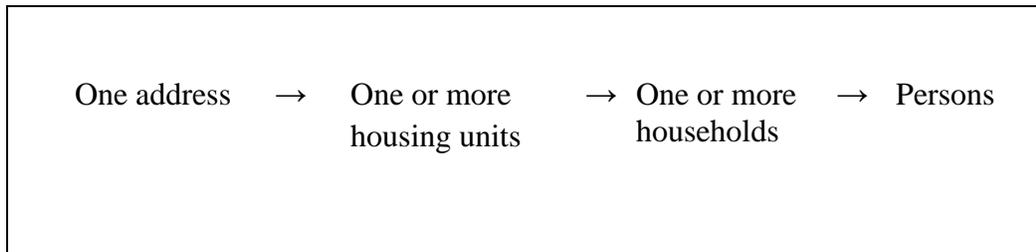
### **Second-stage sample design (from 2006 through 2015)**

This section describes the selection HU and GQ samples within each PSU. The term “living quarters” is used instead of HUs and GQs, unless the distinction is necessary. Complicated relationships may exist between the living quarters on the frame (i.e., addresses) and the living quarters of interest (i.e., HUs and persons). These are deliberately avoided in the descriptions of the second-stage sample design to keep the descriptions simple and readable. An address is a location for mailing and can be associated with multiple HUs or one person because each person could have his or her own address (e.g., PO Box). A HU was defined previously under Coverage. The Census Bureau's definition of household consists of all the persons who occupy a HU. A household includes related family members and all of the unrelated persons, if any, such as lodgers, foster children, wards, or employees who share the HU. A person living alone in a HU or a group of unrelated persons sharing a HU, such as partners or roomers, is also counted as a household. Therefore, a household may include one or more families.

The living quarters of interest for NCVS are the HUs and GQs in the United States and sometimes the persons within the HUs and GQs. However, no simple list of HUs or GQs exists to be used as a frame. With the 2000 design, lists of addresses associated with the HUs and GQs are used. These lists are not comprehensive. For example, many addresses are ineligible because they are associated with businesses and vacant units. Also, a single address may have a complex

association with multiple households, as represented in figure 2.3. An example of this could be a single address for a house that has an entrance to the main level and upstairs in the front and a second entrance to the basement. The landlord lives on the main level; a tenant lives on the upper level, where a hallway is shared; and a third tenant lives in the basement. In this example, one address is comprised of two housing units, two households, and three persons.

**Figure 2.3. Association of addresses with households and persons**



Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2013.

The rest of this section is divided into two parts. In the first part, a description is provided of how the units are selected for any frame with the general methodology of systematic random sample from an ordered list. The second part of the section describes the sample selection for each of the specific frames used in each sample.

***Part I. Systematic random sampling from an ordered list***

For all four frames, a systematic sample is selected from an ordered list (*sys*). As explained by Cochran (1977, p. 208), the *sys* sample design can be more efficient in terms of sampling variances than a simple random sample with replacement when the variable used to sort the list is associated with the variable of interest.

The *sys* sample design starts with a set of known sampling intervals referred to as "take every" (TE), which are the inverse of the selection probability within a PSU. Next, a random start (RS) is calculated as—

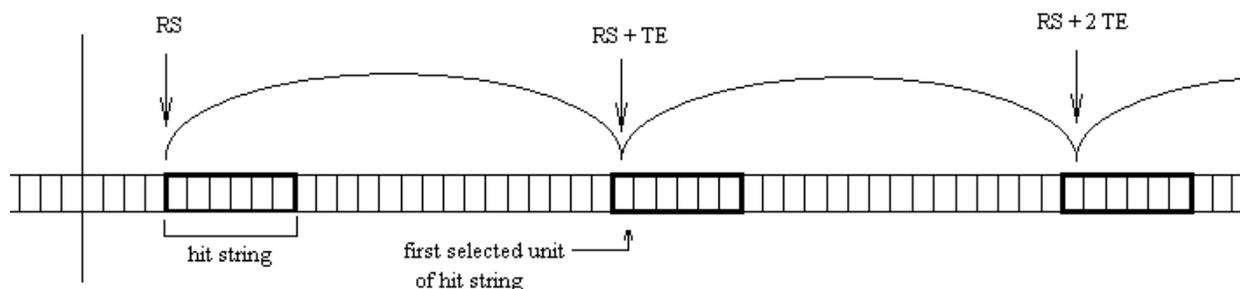
$$RS = RN \times TE$$

where RN is a random number from a uniform distribution on the interval (0, 1]. The RSs and TEs are used to determine the selected units of the sample. The sample includes all of those units from the ordered list corresponding to—

$$RS, RS + ( 1 \times TE ), RS + ( 2 \times TE ), \dots, RS + ( k \times TE )$$

where *k* is the largest integer such that  $RS + ( k \times TE ) \leq N$ , and *N* is the number of units on the frame. Each non-integer number of the sequence will be rounded up to the next largest integer. Operationally, selected units as hits and the set of selected units are referred to as the hit string.

**Figure 2.4. Representation of systematic sampling**



Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2013.

Figure 2.4 illustrates the *sys* sample design. The vertical line on the left side of the figure represents the boundary between the two strata. Three hit strings are selected into the sample. The first hit string starts at RS and includes the next six units. The second and third hit strings start at  $RS + TE$  and  $RS + 2 TE$ , respectively.

### **Efficient rotation of units**

NCVS selected a hit string of 28 housing units for the 10-year sample. The 28 units of the hit string were divided into seven clusters of four units, interviewed throughout the 10-year period of the sample. In the panel design of NCVS, when the four units of the cluster have completed seven interviews, a different cluster of four units from the same hit string rotates into the sample to replace them.

The rotation of clusters from the same hit string is efficient because, given the sort order, the replacement cluster is reasonably similar to the units it is replacing due to the geographic proximity of the units. The expectation is that, when units exit, the entering units will be similar due to correlation. Because units within a hit string are highly correlated, the variance of the estimate is reduced. Therefore, this rotating structure is more efficient than replacing each cluster of four units with a completely random cluster of units.

### **Part II. Frames for the second-stage sample design**

The frame and sampling methodology for the second-stage differ by block within each sample PSU. By block, either the unit/permit/GQ frame or the area frame is used. If the block has complete addresses in the unit frame and a permit issuing office exists (i.e., the permit frame is viable), then the unit/permit/GQ frames are used. If the block does not have complete addresses or no permit issuing office is covering the block, then the area frame is used.

### **Sample selection for the unit frame**

The unit frame is the simplest of all the frames and has simple methods for sample selection. The unit frame is a list of housing units compiled during the previous decennial census. For the 2000 sample design, the frame includes all units constructed before April 1, 2000. For the unit frame, the *sys* sample design sorts the frame with respect to the following variables:

- central city/balance/urban/rural
- county
- tract
- basic street address

- unit sort order.

Once the units of the unit frame are sorted, a *sys* sample of units is selected within each PSU.

### **Sample selection for the permit frame**

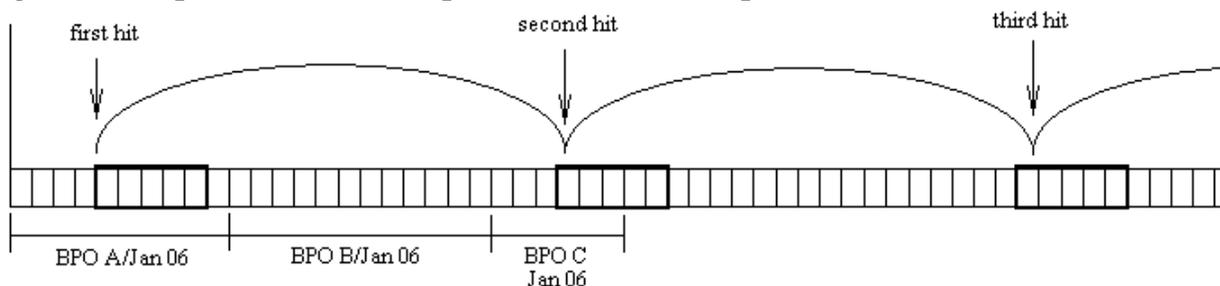
The permit frame represents all housing units constructed after April 1, 2000. However, when selecting the sample of units from the permit frame, it is not possible to identify all of the units that will be constructed in the future. The permit sample is selected shortly after the decennial census and is used for the next 10 years.

The selection of the permit sample involves building a “skeleton frame” of units expected to be constructed in the next 10 years. The units of the frame are a placeholder for the units that will be constructed in the future. As new units are identified from the Building Permit Survey, they are filled into the spots on the skeleton frame, and then a *sys* sample of hit strings from the permit frame is selected. Unlike the other three frames, the units on the permit frame are not sorted before selecting the sample of hit strings. The sort is determined by the process used to add permit units to the skeleton frame. This process is explained below.

It is cost prohibitive to visit every Building Permit Office (BPO) and list building permits in each sampled PSU. In response, the Census Bureau uses results of the Survey of Construction (SOC) to determine how many building permits were issued. The SOC is a jointly conducted survey by the Census Bureau in conjunction with the U.S. Department of Housing and Urban Development. The SOC receives monthly or annual counts of building permits issued for different building permit offices across the country, though these counts differ by frequency, either monthly or annual, depending on the frequency of a given BPO's reporting.

The NCVS regularly matches the SOC counts to the permit frame. The results from the SOC are converted to groups of four housing unit equivalents (usually adjacent or neighboring), called “measures,” because the census addresses of individual housing units or people within a group quarter are not used in the sampling. These measures are continuously accumulated and linked with the frame of hypothetical measures used to select the NCVS sample. This matching identifies which BPO contains the measure that is in the sample. Using an automated instrument, a field representative visits the BPO to obtain a list of 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. From the list of permits, the final sample of newly constructed units is identified.

**Figure 2.5. Representation of sample selection from the permit frame**



Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2013.

Figure 2.5 illustrates the sample selection methodology for the permit frame. Three different sample hit strings have been identified by the sample hit strings. The first hit or first unit of the hit string is identified with an arrow and the other units of the hit string are bold.

Three BPOs have been matched to the skeleton frame in January 2006: 10 units are expected for BPO A, 12 units are expected with BPO B, and six units are expected with BPO C. With the BPOs applied to the permit frame, it is known that five units from BPO A are associated with the first hit string and three units from BPO C are associated with the second hit string.

The two BPOs associated with the sample units are assigned field staff who then visit the BPO to list the actual permits represented by the count. As subsequent BPOs are added to the skeleton frame, the second and third hits identify the additional sample units that require listing by field staff. Because the Census Bureau used the skeleton frame, BPO B is not listed by the field staff based on the January 2006 report because it was not selected for the sample.

### **Sample selection for the GQ frame**

The 2000 decennial census generates a list of the GQs in the United States and the count of persons living in the GQs, but the census does not provide a list of rooms or other types of divisions with the GQs that may be used as a frame. For example, although the number of students living in Citizen Dormitory at Public University is known, a listing of each dorm room in Citizen Dormitory is not known. A sampling methodology similar to the method used for the permit frame is used to select dorm rooms and students within the selected dorm rooms because it is not necessary to interview each student in every dorm room.

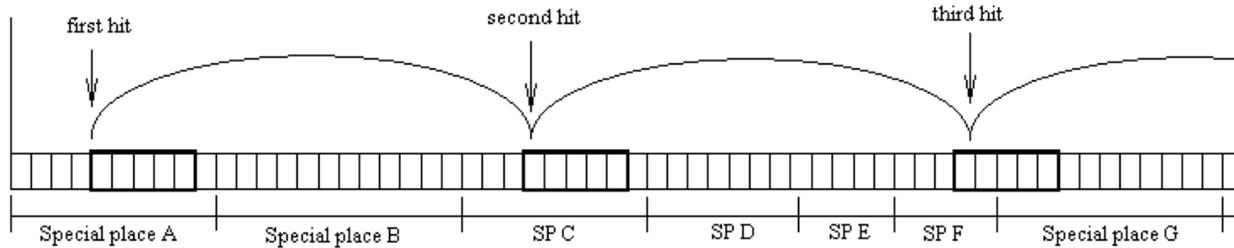
This is accomplished by generating a frame of expected units within each GQ special place (i.e., a single GQ or multiple GQs in the same facility). Each unit of the GQ frame has approximately the same number of persons as the unit frame. For the 2000 design, an average of 2.6 persons per household is used. Therefore, as an example, if the decennial census reported that a given special place had 52 persons, 20 housing unit equivalents are created for the GQ.

As with the other frames, a *sys* sample is selected from the frame after the units of the GQ frame are sorted by the following variables:

- basic primary sampling unit component
- county
- tract
- block
- within block measure number.

Once the sample is selected, it is then possible to identify the GQs in the sample. The sample GQs are listed and the final sample is selected from the list of rooms or other divisions.

**Figure 2.6. Representation of sample selection with the GQ frame**



Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2013.

Figure 2.6 illustrates the sample selection methodology for the GQ frame. The sample selection has identified three different sample hit strings. The first hit or first unit of the hit string is identified with an arrow, and the other units of the hit string are bold. The frame is composed of units, one for each expected 2.6 persons of a special place.

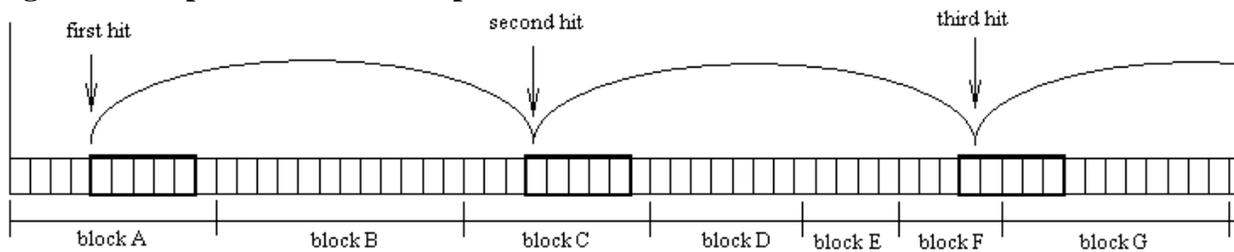
After selecting the sample, special places that are included in sample units are identified. In figure 2.6, special places A, C, F, and G are associated with sample units. Therefore, field staff are sent to list all of the GQs within those blocks. A final sample from the listing is identified.

### Sample selection for the area frame

The selection methodology of the area frame is similar to the permit frame. It ensures that field staff do not list every block in every sample PSU. This is accomplished by generating a frame of expected units. The area frame is initially populated with the expected block counts because only a list of sample blocks is needed. Similar to the permit frame, blocks listed have at least one unit selected into the sample. These units of the area frame are sorted by the following variables:

- basic primary sampling unit component
- state
- combined block central city/balance/urban/rural
- county
- tract
- block.

**Figure 2.7. Representation of sample selection with the area frame**



Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2013.

Figure 2.7 illustrates the sample selection methodology for the area frame. The sample selection has identified three different sample hit strings. To keep the example simple, only five units are selected for each hit string. The first hit or first unit of the hit string is identified with an arrow and the other units of the hit string are bold. The frame is comprised of units, one for each expected unit of the block.

After selecting the sample, the blocks that include sample units and the blocks that exclude sample units are identified. In figure 2.7, blocks A, C, F, and G are associated with sample units. Therefore field staff are sent to list all of the units within those blocks. The final sample from the listing is identified.

### **Subsampling**

For both the permit and the area frame, the expected number of units and the actual number of units listed is often different. When the expected number is different from the actual number, a subsampling method to identify the final sample is used. Additionally, when an interviewer visits the units to conduct the interview, sometimes the interviewer find additional units. Further subsampling is used in this case to maintain reasonable interviewer workloads. Both types of subsampling use the *sys* sample design and both are accounted for in the weighting with the weighting control factor.

### ***Contribution of each stage of sampling to the variance***

The two stages of the sample design are sources of variability for NCVS estimates. By selecting a different sample of PSUs or a different set of households within the sample PSUs, estimates would vary. In subsequent sections, the estimation of overall variance is discussed. The contribution of each stage of the sample design to the overall variance for select statistics is presented below.

Table 2.3 presents coefficients of variation for each year of total personal and property crimes from 2008 through 2013, with percentage of variance accounted for by the first stage and second state of sampling.

**Table 2.3. Percent of variance by stage of NCVS sample design, 2008–2013**

Year	Total personal crime			Total property crime		
	Overall CV	1 <sup>st</sup> stage	2 <sup>nd</sup> stage	Overall CV	1 <sup>st</sup> stage	2 <sup>nd</sup> stage
2008	0.045	17%	83%	0.024	36%	64%
2009	0.040	0%	100%	0.021	22%	78%
2010	0.047	30%	70%	0.019	0%	100%
2011	0.043	7%	93%	0.026	45%	55%
2012	0.043	25%	75%	0.020	9%	91%
2013	0.062	32%	68%	0.023	24%	76%

Note: The coefficient of variation (CV) is the ratio of the standard error to the estimate. CVs provide a measure of reliability and a means to compare the precision of estimates across measures with differing levels or metrics.

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2008–2013.

## Sample sizes

### *Households*

The 2013 survey had a national sample of approximately 66,000 designated addresses located in approximately 319 primary sampling units (self-representing and non self-representing) throughout the United States. The discrepancy between the 66,000 mentioned here and the half-year sample size is due to the panel design, with some housing units rotating out of the sample and new housing units rotating into the sample during the entire year.

Table 2.4 shows a distribution of the unique housing units sampled for calendar year 2013 by the Core-Based Statistical Area (CBSA) status and location of residence. CBSA status is determined by the Office of Management and Budget (OMB), and location of residence (i.e., urban, suburban, and rural) is determined by the Census Bureau for all of the current surveys, including NCVS (see Terms and Definitions at the end of this report). Note that different techniques were used to calculate the sample size, as seen previously in this section (i.e., half-year), and the number of housing units calculated here and during re-instatement (i.e., calendar year), which explains why the counts are not in exact agreement.

**Table 2.4: Distribution of NCVS housing units sample, by CBSA status and location of residence, 2013**

<b>CBSA status</b>	<b>Location of residence</b>	<b>Housing units</b>	<b>Percent</b>
Central city of a CBSA	Urban	21,091	32.1%
Central city of a CBSA	Rural	1,276	1.9%
Balance of CBSA	Suburban	29,221	44.5%
Balance of CBSA	Rural	4,165	6.3%
Not in a CBSA	Rural	9,852	15.0%
<b>Total</b>		<b>65,605</b>	<b>99.8%</b>

Note: A Core-Based Statistical Area (CBSA) is a geographic area consisting of the county or counties or equivalent entities associated with at least one core (urbanized area or urban cluster) of at least 10,000 population, plus adjacent counties having a high degree of social and economic integration with the core as measured through commuting ties with the counties associated with the core. Percentages do not sum to 100% due to rounding.

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2013.

### *Persons*

For calendar year 2013, NCVS interviewed (i.e., final sample size) about 106,000 eligible persons age 12 or older. Table 2.5 shows a distribution of persons by sex and race (black/African American and non-black/African American). This distribution may be useful to check whether any sample expansion or reduction have similar composition or not.

**Table 2.5. Distribution of persons in NCVS sample, by sex and race, 2013**

<b>Sex</b>	<b>Race</b>	<b>Number of persons</b>	<b>Percent</b>
Male	Black/African American only	5,216	4.9%
Male	Non-black/Non-African American	45,064	42.5%
Female	Black/African American only	6,730	6.4%
Female	Non-black/Non-African American	48,952	46.2%
<b>Total</b>		<b>105,962</b>	<b>100%</b>

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2013.

### *Sample reinstatement*

Based on budgetary decisions at the Bureau of Justice Statistics (BJS), sample reductions began in 2006 and reduced the sample to about 51,000 unique/unduplicated housing units at the end of 2009. Beginning in October 2010, cases from prior sample reductions were reinstated to increase the sample size by 20% to approximately 64,000 cases per year. Table 2.6 presents the number of unique (i.e., unduplicated) cases reinstated by quarter from 2010 through 2013 for each of the four frames.

**Table 2.6. Number of unduplicated NCVS housing units reinstated by quarter and frame, 2010–2013**

Year - Quarter	Frame				
	Area	GQ	Permit	Unit	Total
2010 - Quarter 4	0	0	537	4,284	4,821
2011 - Quarter 1	547	17	723	4,299	5,586
2011 - Quarter 2	711	10	106	869	1,696
2011 - Quarter 3	94	0	170	741	1,005
2011 - Quarter 4	69	0	77	507	653
2012 - Quarter 1	155	2	168	1,080	1,405
2012 - Quarter 2	182	3	166	1,335	1,686
2012 - Quarter 3	127	0	182	961	1,270
2012 - Quarter 4	155	0	135	991	1,281
2013 - Quarter 1	142	1	164	953	1,260
2013 - Quarter 2	156	0	131	912	1,199
2013 - Quarter 3	127	6	116	922	1,171
2013 - Quarter 4	147	0	186	927	1,260
Total 2010 only	0	0	537	4,284	4,821
Total 2011 only	1,421	27	1,076	6,416	8,940
Total 2012 only	619	5	651	4,367	5,642
Total 2013 only	572	7	597	3,714	4,890
Total 2010–2013	2,612	39	2,861	18,781	24,293

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2010–2013.

Table 2.7 presents the distribution of unique housing units reinstated in calendar year 2013 by CBSA status and location of residence. This distribution reveals a similar sample composition to the total unique housing units for the entire sample presented in table 2.4.

**Table 2.7. Distribution of NCVS housing units from reinstated sample, by CBSA status and location of residence, CY 2013**

<b>CBSA status</b>	<b>Location of residence</b>	<b>Housing units</b>	<b>Percent</b>
Central city of a CBSA	Urban	1,364	37.0%
Central city of a CBSA	Rural	71	1.9%
Balance of CBSA	Suburban	1,477	40.1%
Balance of CBSA	Rural	222	6.0%
Not in a CBSA	Rural	550	14.9%
<b>Total</b>		<b>3,684</b>	<b>99.9%</b>

Note: A Core-Based Statistical Area (CBSA) is a geographic area consisting of the county or counties or equivalent entities associated with at least one core (urbanized area or urban cluster) of at least 10,000 population plus adjacent counties having a high degree of social and economic integration with the core as measured through commuting ties with the counties associated with the core. Percentages do not sum to 100% due to rounding.

Source: Bureau of Justice Statistics, National Crime Victimization Survey, CY 2013.

Table 2.8 shows that the distribution of persons in reinstated housing units by sex and race (black/African American and non-black/African American) has a similar composition to the whole sample during calendar year 2013, as seen in table 2.5.

**Table 2.8. Distribution of NCVS persons from reinstated sample, by sex and race, CY 2013**

<b>Sex</b>	<b>Race</b>	<b>Persons</b>	<b>Percent</b>
Male	Black/African American only	1,003	4.9%
Male	Non-black/Non-African American	8,719	42.4%
Female	Black/African American only	1,348	6.6%
Female	Non-black/Non-African American	9,493	46.2%
<b>Total</b>		<b>20,563</b>	<b>100%</b>

Source: Bureau of Justice Statistics, National Crime Victimization Survey, CY 2013.

## Chapter 3. Data collection

### Field operations

#### *Field representative training*

Training for NCVS interviewers consists of classroom and on-the-job training. Initial training for field representatives consists of a full day pre-classroom self-study, 4-day classroom training, post-classroom self-study, and on-the-job observation and training.

Initial training includes topics such as probing for accuracy, the power of persuasion, case management, Automated Listing and Mapping Instruments (ALMI), NCVS concepts and definitions, Computer-Assisted Personal Interviewing (CAPI) instrumentation training and exercises, protocols for transmission of work, and topics related to performance measures. Experienced NCVS FRs are provided with ongoing basic refresher training that uses the following methods: refresher self-studies that cover NCVS procedures or concepts that may generate interviewing errors, monthly memoranda that address ongoing operational issues and occasionally include review exercises to reinforce survey concepts and procedures, and ongoing feedback from observations of interviews by supervisors or senior field representatives. At one time, refresher training program also included routine classroom training. However, this component of the program was suspended in the 1990s. Classroom refresher training for NCVS FRs was reinstated toward the end of 2011 and the beginning of 2012.

The Census Bureau employs about 750 FRs across the United States to administer NCVS interviews. An FR refresher training program was developed in the summer of 2011 and implemented in August 2011, which included learning detailed information on the survey, screener questions, crime incident reports and concepts, and contact history instrument; fielding respondent questions; asking sensitive questions; and gaining respondent cooperation. This training also focused on the NCVS CAPI with a renewed focus on data quality, including a training workbook and paired practice interviews. Data quality indicator reports were also discussed and reviewed.

The refresher training aimed to reacquaint FRs with the purpose and content of the screen questions (NCVS-1) used to identify whether a respondent suffered a victimization. It also intended to clarify the information on the incident follow-up form (NCVS-2), which collects details about the characteristics of each incident. Because crime is a relatively infrequent event, many FRs conduct a large number of interviews without uncovering a criminal event. Therefore, FRs must maintain familiarity with the questionnaire to conduct a proper interview when a respondent reports a crime. In addition, it is important that FRs across every regional office conduct the interview in a standardized manner to ensure a high quality survey.

Along with FR refresher training, the Census Bureau implemented a series of field supervisory performance and data quality measures. Previously, high response rates were the primary measure of FR performance. Under the revised performance structure, FRs are monitored on response rates (i.e., household and person), screener time stamps (i.e., the time it takes to administer the screener questions on the NCVS-1 instrument), early and overnight interview starts (i.e., interviews conducted very late in the evening or very early in the morning), contact history with household (i.e., number of attempts to contact the household), and completeness of screen and incident instruments (i.e., whether items are missing). Any noncompliance with these measures led to supervisor notification and follow-up with the FR. The follow up activity may

include simple points of clarification (e.g., the respondent works nights and is only available in the early morning for an interview), additional FR training, or FR removal from the survey.

### *Effects of refresher training*

To measure impacts on key variables and performance, a phased-in experimental design was used to implement the refresher training project. Teams of FRs were randomly assigned to two cohorts, with Cohort 1 receiving the intervention first (i.e., refresher training and field performance monitoring) and Cohort 2 serving as the control group. Cohort 1 received training starting in August 2011, and field performance measures and monitoring began in October 2011. Cohort 2 did not receive any refresher training or any additional field monitoring at the time. Starting in January 2012, Cohort 2 received refresher training with the additional performance measures and field monitoring. By March 2012, the majority of all FRs administering the NCVS collection had received refresher training and were under the new field monitoring and performance system.

For both cohorts, along with the new performance measures, key variables of interest were the number and type of crime incidents collected per cases worked. Crime counts per cohort were monitored 3 months before the August refresher training to serve as a pre-test baseline measure. These counts were monitored throughout the experimental design. Overall, the experimental design allowed the Census Bureau and BJS to randomly assign FR teams to cohorts, account for any pre-existing differences in crime incident counts per sample caseload, compare Cohort 1 and Cohort 2 from August 2011 through January 2012, and then continue to monitor any differences between cohorts throughout 2012 after both were trained and under the new performance management system. The Cohort 1 cases used in the experimental design for training were not used to generate estimates for 2011. The Census Bureau and BJS continued to monitor the cohorts in 2012. Refresher training was a major intervention for the 2011 and 2012 survey years and its effect on the crime incidence collection was heavily reviewed and analyzed.

A comprehensive evaluation of the 2011 experiment is available on the Census Bureau's website <<https://www.census.gov/srd/papers/pdf/rrs2013-07.pdf>>.

### *Contact and follow-up procedures*

The initial interview with a sample household is always conducted by personal visit. An initial interview is conducted with the most knowledgeable household member regarding property crimes affecting the entire household. If it is not possible to obtain face-to-face interviews with all other eligible members of the household during the initial interview contact, interviews by telephone are permissible thereafter. NCVS interviews conducted in subsequent enumeration periods are generally conducted by telephone.

The actual time required to interview all eligible members of a sample household varies depending on the household's composition and crime experiences during the reference period. To elicit more accurate reporting of incidents, NCVS uses the self-respondent method, which requires the direct interview of each person age 12 or older in the household.

The primary instruments used for data collection include the NCVS control card and two NCVS questionnaires, the NCVS-1 and the NCVS-2. (See the survey instruments section below for more information.) The control card is the basic record for each sample unit throughout each enumeration period. The NCVS-1, the Basic Screen Questionnaire, contains questions designed to determine whether any crimes were committed against the household as a whole or against an

individual household member during the 6-month reference period. The NCVS-2, the Crime Incident Report (CIR), is used to gather detailed information about crimes reported in the Basic Screen Questionnaire. The interviews conducted in the field are done with a CAPI instrument.

In 2011, the NCVS interview time for respondents reporting no crime incidents averaged 4.95 minutes; the average time for respondents reporting one or more incidents was 21.76 minutes. NCVS interview time increased slightly in 2012, as the time for respondents reporting no crime incidents averaged 7.47 minutes, and the average time for respondents reporting one or more incidents was 24.33 minutes. For 2013, the NCVS interview for respondents reporting no crime incidents averaged 5.98 minutes; the average for respondents reporting one or more incidents was 22.76 minutes. The actual time required to interview each eligible member of a sample household varies depending on the household's composition and crime experiences during the reference period.

## Survey instruments

### *Overview*

The NCVS data collection process is conducted in three general stages: (1) generate roster and identify household respondents, (2) screen for potential victimizations, and (3) classify and date victimizations.

### *NCVS-500 Control Card/household roster and demographic characteristics*

The NCVS-500 Control Card is the basic record for each sample unit throughout each enumeration period and is used by FRs to locate and confirm that they have contacted the correct sample household. The control card contains the sample unit's control number, address, and basic information about the sample household, including the name, age, sex, race and ethnicity, marital status, and education level of each person living in the household. The control card also provides information about the housing unit, the household's total income, the number of contacts made with the household, and information on noninterviews.

The NCVS-500 Control Card may be found on the BJS webpage dedicated to the NCVS: [www.bjs.gov](http://www.bjs.gov).

### *NCVS-1 Basic Screen Questionnaire*

The NCVS-1 Basic Screen Questionnaire contains questions designed to determine whether any crimes were committed against the household as a whole or against an individual household member during the 6-month reference period. The NCVS-1 contains sections for the household respondent's interview and additional household member interviews. Questions in this section are written in a "short cue" format, with the interviewer reading a question "stem" about whether the respondent has experienced a certain type of incident, and then giving examples of the type of incident in "short cues" to prompt respondent memory. The interview proceeds on a person-by-person basis until a questionnaire is completed for each person age 12 or older in the household. For entire households that refuse to participate in the survey or are not available during the interview period, an NCVS-7 Noninterview Record is completed, which contains selected information about the household, such as type of housing unit, reason for noninterview, public housing status, whether the unit is in Indian country, and whether access to the unit is restricted in any way.

The NCVS-1 Basic Screen Questionnaire may be found on the BJS webpage dedicated to the NCVS: [www.bjs.gov](http://www.bjs.gov).

***NCVS-2 Crime Incident Report/topical sections***

The NCVS-2 Crime Incident Report (CIR) is used to gather detailed information about crimes reported in the NCVS-1 Basic Screen Questionnaire. One NCVS-2 CIR is completed for each incident of crime reported in the NCVS-1. For example, if a respondent said that he/she was robbed once and was later beaten up twice, then three NCVS-2 CIRs are completed—one for the robbery, one for the initial assault, and one for the second assault.

Within the CIR, questions are asked in topical modules: location and presence, attack/threat/injury/medical care, emotional toll, actions against the offender, offender, attempted and completed thefts, property damage and police, activity at the time of incident and time/money lost, series of crimes, hate crime, disability, and a written summary of the incident.

The NCVS-2 Crime Incident Report (CIR) may be found on the BJS webpage dedicated to the NCVS: [www.bjs.gov](http://www.bjs.gov).

***Supplements***

Questions or “supplements” are regularly added to the end of the NCVS interview to make timely estimates of specific types of victimization. Supplement instruments are administered to all of those eligible for that particular supplement directly after asking the NCVS-1 and NCVS-2 questions. Eligibility for the supplement varies depending on which supplement is being fielded.

Table 3.1 lists some of the most recent supplements. The two supplements for the 2011 survey year were the School Crime Supplement (SCS), which was fielded from January through June, and the Police–Public Contact Supplement (PPCS), which was fielded from July through December. In 2012, the Identity Theft Supplement (ITS) was fielded from July through December. In 2013, the SCS was fielded from January through June.

**Table 3.1. NCVS supplements, by name and date conducted, 1989–2013**

<b>Supplement name</b>	<b>Dates conducted</b>	<b>Universe of interest</b>	<b>Topic</b>
Police–Public Contact Survey	2011, 2008, 2005, 2002, 1999, 1996	NCVS respondents age 16 or older	Prevalence of contact with police
School Crime Supplement	2013, 2011, 2009, 2007, 2005, 2003, 2001, 1999, 1995, 1989	NCVS respondents ages 12–18 enrolled in school	School-related victimization
Identity Theft Supplement	2012, 2008	NCVS respondents age 16 or older	Identity theft
Workplace Risk Supplement	2002	NCVS respondents age 16 or older currently employed or employed at least 2 weeks in the last 6 months	Workplace violence
Supplemental	2006	NCVS respondents	Harassment or

Victimization Survey (Stalking)	age 18 or older	unwanted contact/behavior
------------------------------------	-----------------	------------------------------

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 1989–2013.

### **School Crime Supplement, 2011 and 2013**

Cosponsored by the National Center for Education Statistics (NCES) and BJS, the SCS collects information about victimization, crime, and safety at school. The SCS surveys a nationally representative sample of approximately 6,500 students age 12 to 18 who attend public and private elementary, middle, and high schools in the United States. The SCS was conducted in 1989, 1995, 1999, 2001, 2003, 2005, 2007, 2009, 2011, and 2013. The supplement inquires about school-related topics, such as—

- alcohol and drug availability
- fighting, bullying, and hate-related behaviors
- fear and avoidance behaviors
- gun and weapon carrying
- gangs at school.

No substantial changes were made to the 2013 SCS questionnaire. In 2011, the SCS questionnaire was revised to include new questions that asked about perceptions of crime and safety in the respondent’s neighborhood and the neighborhood in which the respondent’s school is located. In addition, two questions were modified. The question about relationships and social bonds with adults at school was reworded and expanded with new response categories. These categories included whether students agreed that there was an adult at school who really cares about them, notices when they are not there, listens to them, tells them when they do a good job, always wants them to do their best, and believes that they will be a success. Also, a category was added to the cyber-bullying questions that asked whether another student purposely shared the respondent’s private information, photos, or videos on the Internet or on a mobile phone in a hurtful way. Administratively, refusals to participate were separated into two distinct categories so program managers could distinguish respondent self-refusals from those by a parent or other gatekeeper.

Approximately 9,500 sample persons were eligible for the 2013 SCS supplement; approximately 5,700 of these persons were successfully interviewed—a response rate of 60.0%.

### **Police–Public Contact Survey, 2011**

The PPCS provides detailed information on the nature and characteristics of face-to-face contacts between police and the public, including the reason for and outcome of the contact, and the respondent’s satisfaction with the contact. The PPCS interviews a nationally representative sample of more than 60,000 residents age 16 or older about any voluntary or involuntary contacts with police during the previous 12 months. The PPCS was conducted in 1996, 1999, 2002, 2005, 2008, and 2011. The survey enables BJS to estimate the likelihood of a driver being pulled over in a traffic stop and the percentage of all contacts that involve the use of force by police.

In 2011, major revisions were made to the screener section of the PPCS. Before 2011, the PPCS screener consisted of two questions. The first question asked whether the respondent had any

face-to-face contacts with police in the previous 12 months. If the respondent reported any face-to-face contacts with police, a second screener question asked how many face-to-face contacts the respondent had with police. In 2011, the PPCS screener was restructured to ask about each specific type of contact with the police. This approach is expected to elicit more reports of contact with police, regardless of whether the contact was face-to-face or not. Before 2011, all contacts were categorized as either traffic stops or non-traffic stops. Follow-up questions on reason for contact, characteristics and outcome of the contact, and whether police used force during the contact were also asked. In 2011, contacts were categorized into street stops, traffic stops, and other voluntary contacts. Similar follow-up questions were asked on the reason for contact, characteristics and outcome of the contact, and whether police used force during the contact. Due to the significant changes made to the 2011 version of the questionnaire, BJS asked that the Census Bureau field both the 2011 and 2008 PPCS questionnaires in a split sample test to compare the contact rates collected in the two versions. Approximately 85% of respondents received the 2011 PPCS questionnaire, and 15% received the 2008 PPCS questionnaire.

Approximately 62,000 respondents were eligible for the 2011 PPCS supplement; approximately 49,000 of these persons were interviewed successfully—a response rate of 79.1%.

### **Identity Theft Supplement, 2012**

The ITS generates overall estimates of identity theft and demographic characteristics of victims who reported one of the following types of identity theft:

- Unauthorized use or attempted use of an existing account, such as a credit/debit card, checking, savings, telephone, online, or insurance account.
- Unauthorized use or attempted use of personal information to open a new account, such as a credit/debit card, telephone, checking, savings, loan, or mortgage account.
- Misuse of personal information for a fraudulent purpose, such as getting medical care, a job, or government benefits; renting an apartment or house; or providing false information to law enforcement.

The ITS details the victims' direct and indirect financial losses; the time spent resolving problems related to the identity theft; the percentage of victims who reported the theft to credit card companies, credit bureaus, and law enforcement agencies; and the level of distress felt by identity theft victims. The ITS was previously first fielded in 2008. However, the two supplements are not directly comparable due to methodological differences

The 2008 ITS was designed to separate victims of attempted identity theft from victims of successful identity theft by screening for monetary loss. However, post-collection analyses demonstrated that this screening approach resulted in an unnecessarily complicated instrument that was not able to clearly distinguish between attempted and completed incidents. Ultimately, the decision was made not to distinguish between completed and successful identity thefts in the screener. The 2012 ITS was refined accordingly, thereby enabling similar questions to be asked of all victims, without a distinction between attempted and completed identity theft.

Classification of attempted and completed incidents was established during post-collection analysis and was based on the amount of monetary loss sustained by a victim.

In addition, the 2008 ITS used a 2-year reference period, which was shown to inhibit respondent recall because respondents had difficulty remembering events that occurred more than one year

prior. The 2012 ITS used a 1-year reference period and included a new section on the long-term consequences of identity theft to measure victims who experienced identity theft more than one year prior and continued to deal with the consequences. In keeping with the definition of identity theft used in the 2008 ITS, the 2012 ITS asked respondents to report if they experienced the misuse of an existing credit card, misuse of another existing account (e.g., checking and savings), misuse of personal information to open a new account, or misuse of personal information for other fraudulent purposes in the 12 months prior to interview.

Approximately 87,000 respondents were eligible for the 2012 supplement; approximately 64,000 of these persons were interviewed successfully—a response rate of 73.8%.

## Interview procedures

### *Consent and confidentiality*

All of the data for NCVS are collected by BJS under the authority of Title 42 U.S.C. § 3732. In addition, BJS is required to keep all personally identifying information about respondents strictly confidential, under the authority of Title 42 U.S.C. §§ 3789g and 3735.

All information collected as part of the NCVS is held in strictest confidence under Title 13 of the United States Code and is seen only by sworn employees or agents of the Census Bureau. Any information from the survey that is disclosed or released to others will be handled in such a manner that persons cannot be identified. Unauthorized disclosure of individual information by a sworn Census Bureau employee is punishable by a fine of up to \$250,000, imprisonment of up to 5 years, or both.

The confidentiality statement on the NCVS questionnaire reads:

*“We are conducting this survey under the authority of Title 13, United States Code, Section 8. Section 9 of this law requires us to keep all information about you and your household strictly **confidential**. We may use this information only for statistical purposes. Also, Title 42, Section 3732, United States Code, authorizes the Bureau of Justice Statistics, Department of Justice, to collect information using this survey. Title 42, Sections 3789g and 3735, United States Code, also requires us to keep all information about you and your household strictly confidential.”*

The Privacy Act of 1974 requires federal agencies to provide persons with the following information when collecting personal information:

- authority: Title 13 U.S.C. § 182
- compliance: voluntary
- penalty for not participating: none.

Information collected in the survey is released only in the form of summary statistics. Information concerning a person is never available to anyone except sworn Census Bureau employees. Other government agencies, including the FBI and IRS, cannot gain access to individual Census Bureau records.

Respondents are asked to report crime experiences occurring in the 6 months preceding the month of interview. Research has shown that respondents are able to recall events more accurately over a 6-month period than over a 12-month period, and recall an event that occurred within 3 months of the interview more accurately than one that occurred within 6 months.

However, a shorter reference period would require more interviews per year, significantly increasing data collection costs. These increased costs would have to be balanced by cost reductions elsewhere, and sample size is often considered. However, reducing sample size diminishes the precision of estimates of relatively rare crimes. In light of these trade-offs of cost and precision, a reference period of 6 months is used for the NCVS, and some degree of response error is accepted.

All of the eligible persons are asked if they were victims of crimes that occurred within the previous 6 months. Incidents reported in the NCVS are associated with the address of the respondents and not associated with where the incident occurred.

Interviewers have notes from the previous interview that can be used to check duplicate responses. For example, if persons say that they experienced a robbery in the previous 6 months, and in their previous interview they said that they were robbed, the interviewer confirms that the reported robbery is not the same as previously reported. Overreporting due to reporting incidents outside the reference period is called telescoping (Neter & Waksburg, 1964), and using reported incidents from the prior interview to confirm duplicate reports is called bounding.

## Modes

### *In-person/telephone*

Most NCVS interviews are attempted by telephone because it is more cost effective. An NCVS interview should be conducted in-person when the sample household—

- is assigned for a first enumeration period interview
- has not been interviewed in any previous enumeration period
- does not have a telephone on which they can be reached
- does not want to be interviewed by telephone
- has a privacy detector that requires the caller to enter a personal identification number (PIN).

Table 3.2 shows the percentages of personal visit and telephone interviews by year.

**Table 3.2 NCVS interview mode, by year, 2011, 2012, and 2013**

Year	Personal visit	Telephone
2013	45%	55%
2012	45%	55%
2011	46%	54%

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2011, 2012, and 2013.

### *Use of proxy interviews*

A proxy interview is one in which someone other than the intended household member answers the interview questions for another eligible household member. The person who is authorized to answer for the intended household member is referred to as the proxy respondent. The intended household member who is unable to answer the interview for himself/herself is referred to as the

proxy person. Because a proxy respondent is more likely to omit an incident or leave out some of the details about a reported incident, proxy interviews are discouraged, except as a last resort. Exceptions may be made during data collection to use proxy interviewing instead of direct interviewing in three circumstances:

1. A parent or guardian refuses to allow the interviewer to interview his/her child age 12 or 13. In this case, any knowledgeable household member who is at least age 18 may be the proxy respondent for the child.
2. A household member who is age 12 or older is temporarily absent and will not be back to the address until after the interview closeout date. In this case, the interviewer must make sure that the person is still a household member. If the absent person is still considered a household member, then a proxy interview may be accepted.
3. If a household member who is at least age 12 is considered physically and/or mentally incapacitated, then a proxy interview may be accepted. To qualify as physically and/or mentally incapacitated, the household member must have health and/or mental illness problems that are continuous throughout the entire interview period, and these problems make it impossible for the person to be interviewed.

The following problems do not qualify as health or mental illness problems: colds or the flu, drunkenness, drugs, and emotional problems that might be aggravated due to some NCVS questions, such as those dealing with sexual assaults.

Proxy interviews may not be accepted under the following circumstances:

1. If a household member cannot be reached at the sample address, despite repeated attempts throughout the interview period.
2. If in a two-person household, one of the members says that they cannot take the time to answer the questions and instead wants the other household member to answer all of the NCVS questions for both of them.
3. If a household member refuses to allow an interview of someone in the household who is over age 13.
4. If a respondent does not understand English, and an acceptable interpreter (including household members) cannot be found.

In the above situations, the respondent is classified as a Type Z person noninterview, or person nonresponse, which are noninterviews of persons within a household wherein at least one person, the household respondent, has been interviewed.

From 1973 through 1986, all household members age 14 or older were interviewed directly to determine whether they had experienced any violent or personal property crimes. Proxy interviews were obtained for the youngest eligible respondents, those ages 12 to 13. However, studies conducted as part of the 1976 National Academy of Sciences (NAS) panel, which recommended improvements to the NCVS, indicated that the proxy interview produced less reliable data than direct interviews. Consequently, the NCVS redesign project advocated interviewing all of the respondents age 12 or older—a procedure BJS adopted beginning in July 1986.

Proxy interviews currently account for about 3% to 4% of all NCVS interviews. In 2011, 4.15% of NCVS interviews were conducted by proxy. In 2012, 4.42% were conducted by proxy. In

2013, 4.79% were conducted by proxy. Table 3.3 presents a breakout of proxy interviews by reason.

**Table 3.3. NCVS proxy interviews, by reason, 2011, 2012, and 2013**

	<b>2011</b>	<b>2012</b>	<b>2013</b>
Age 12–13	725	1,371	1,023
Physically/mentally unable	2,442	2,833	3,069
Unavailable	2,759	2,993	3,576
Other	4	8	7

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2011, 2012, and 2013.

### *Replacement households*

The addresses selected for interviewing in the NCVS remain in sample for 3 years, regardless of changes in the household composition. The composition of some households may change due to many reasons, including relocation, marriage, divorce, death, and/or changes in roommates. Regardless of these individual changes in household composition, all of the persons living in the household who are age 12 or older are eligible to be interviewed during each enumeration.

When entire households move away and are replaced by new households, the new households are considered replacement households. Interviewing continues in the usual manner, with eligible members of the replacement household. Replacement households accounted for 3.95% of households in 2011, 4.33% in 2012, and 4.26% in 2013.

### *Spanish and alternative language questionnaires*

Currently, the NCVS questionnaire is available in both English and Spanish. (NCVS introductory letters are also available in Spanish, Chinese, Korean, and Vietnamese.) For Spanish and other languages, field representatives may use an interpreter, if acceptable to the respondent. The interpreter may be a family member, a neighbor of the respondent, an official interpreter, or the field representative if he/she is fluent in the respondent's language.

If finding a suitable interpreter is difficult while making contact, field supervisors may help field representatives obtain assistance. If a suitable interpreter cannot be located, proxy respondents are not allowed for the NCVS. Table 3.4 shows the number and percentage of interviews for the five largest groups of non-English interviews in 2011, 2012, and 2013.

**Table 3.4. To five non-English languages in NCVS interviews, by number and percent, 2011, 2012, and 2013**

Language	2011		2012		2013	
	Total	Percent	Total	Percent	Total	Percent
Total non-English	5,570	3.8%	7,306	4.5%	7,135	4.5%
Spanish	4,735	3.2%	6,204	3.8%	6,090	3.8%
Chinese	252	0.2%	304	0.2%	370	0.2%
Vietnamese	74	0.1%	135	0.1%	109	0.1%
Russian	67	0%	94	0.1%	76	0.1%
Korean	65	0%	87	0.1%	88	0.1%
Other non-English	377	0.3%	482	0.3%	402	0.3%

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2011, 2012, and 2013.

### Interview auditing and reinterviews

The reinterview is a component of the quality control program. It requires that a supervisor or a senior field representative recontact respondents at a sample of households previously interviewed by a regular NCVS field representative. During the reinterview, many questions from the original NCVS interview are re-asked. In 2011 and 2012, about 3.5% of NCVS households were reinterviewed.

The reasons for maintaining a reinterview program are to—

- detect and deter field representatives who may be falsifying data
- provide the supervisory staff with a means of evaluating individual field representatives
- call attention to a need for revising procedures or instructions
- determine whether refresher training is required on some particular phase of the survey.

The reinterview also verifies that the interviewer contacted the correct sample unit, determined the correct household composition, and classified noninterview households correctly. It is also used to estimate the number of missed crimes.

Independence must be maintained between the original interview and the reinterview. To ensure independence, reinterviewers are not allowed to see or edit the original responses or to observe or take part in the preparation of the reinterview materials for households in the reinterview assignment.

To save costs, most reinterviews are conducted by telephone. Personal visits for reinterview are usually only conducted when a valid phone number is not available or the respondent does not want to respond by phone.

Evaluations are conducted to check field representatives' work for coverage and content. In a listing check, the supervisor repeats the listing by recanvassing each designated segment and checking the found living quarters against the list originally prepared by the field representative. In coverage reinterviews, the number of persons in each household is checked against the listings prepared by the field representative. Errors arising out of incorrect listing, failure to conduct

interviews at the correct sample addresses, noninterview misclassifications, and errors found in the roster of persons are recorded.

Reinterview involves verifying the household roster and tenure, and re-asking a subset of the crime screener questions. The original household respondent is the only person eligible to answer the household crime screener questions during reinterview. Only one household member (i.e., the reinterview sample person) age 12 or older will be reinterviewed for the individual screener questions. If the reinterview sample person is the household respondent, the individual screener questions are asked of the household respondent. If the reinterview sample person is someone other than the household respondent, the household respondent is asked only the household screener questions during reinterview and not the individual screener questions.

## **Nonresponse**

As in most surveys, three types of missing data are in the NCVS: household nonresponse, person nonresponse, and item nonresponse. Household nonresponse (i.e., whole unit nonresponse) occurs when an interviewer finds an eligible household's address but obtains no interview. This can happen as a result of a person not being at home or being unwilling or unable to participate in the survey. Household nonresponse is dealt with through weighting adjustments.

Person nonresponse (i.e., Type Z nonresponse) occurs when an interview is obtained from at least one household member, but an interview is not obtained from one or more other eligible persons in that household. Similar to household nonresponse, person nonresponse may happen as a result of a person being unwilling, unable, or unavailable to answer questions. Person nonresponse is dealt with through editing and imputation.

Item nonresponse occurs when a respondent completes part of the questionnaire but does not answer one or more individual questions. Item nonresponse can occur under any of the following circumstances: a respondent refuses or is unable to provide requested information; a response is inconsistent with related responses or is incompatible with response categories; an interviewer fails to ask a question or record an answer; an interviewer makes an error when recording or keying a response; or instrument error results in an item being unasked, skipped, or missed. For item nonresponse, data are generally imputed for core items.

### ***Types of unit nonresponse***

An NCVS sample address can be classified as a noninterview if the living quarters is occupied but the interviewer is unable to obtain a complete interview; if the living quarters is occupied by persons who are not eligible respondents for the NCVS; if the living quarters is vacant; or if the living quarters is not eligible for sample because it is no longer used as a residence, it no longer exists, or it does not qualify based on the current listing and coverage rules. Noninterviews are broken down into three subcategories: Type A, B, and C noninterviews.

Type A noninterviews occur when sample households consist of persons who are eligible for interview, but none of the persons can be interviewed for a specific reason. Examples of this include refusals to participate in the NCVS or instances when no one is home at the sample address.

Type B noninterviews occur when the sample household is not eligible for interview during the current interview period but could become eligible later. Examples of this include vacant sample addresses or households occupied entirely by persons who have a usual residence elsewhere.

Type C noninterviews occur when the sample address should be permanently removed from the NCVS sample. Examples of this include instances in which the housing unit has been demolished or the unit has converted to a permanent business or storage.

Table 3.5 lists all valid outcome codes, including sample household noninterview types, and the percentage of 2011, 2012, and 2013 NCVS cases for each valid outcome code.

<b>Table 3.5. NCVS outcome codes and descriptions, 2011, 2012, and 2013</b>				
<b>Outcome code</b>	<b>Description</b>	<b>Percent of cases</b>		
		<b>2011</b>	<b>2012</b>	<b>2013</b>
201	Completed interview (no Type Zs)	62.71%	59.41%	58.83%
203	Sufficient partial - no more follow up needed	12.31%	13.08%	12.00%
204	Sufficient partial - follow up needed	.01%	.01%	.00%
	<b>Total type A</b>	<b>8.26%</b>	<b>11.24%</b>	<b>13.07%</b>
211	Type A - Duplicate	.02%	.02%	.05%
213	Type A - Language problems	.05%	.05%	.06%
216	Type A - No one home	2.07%	2.41%	2.64%
217	Type A - Temporarily absent	.24%	.31%	.29%
218	Type A - Refused	4.83%	7.10%	8.26%
219	Type A - Other occupied	1.05%	1.35%	1.77%
	<b>Total type B</b>	<b>15.45%</b>	<b>15.27%</b>	<b>15.06%</b>
225	Type B - Temporarily occupied by persons with usual residence elsewhere	1.17%	1.18%	1.20%
226	Type B - Vacant - regular	11.74%	11.55%	11.43%
227	Type B - Vacant - storage of household furniture	.95%	.91%	.87%
228	Type B - Unfit or to be demolished	.41%	.38%	.35%
229	Type B - Under construction, not ready	.30%	.30%	.35%
230	Type B - Converted to a temporary business or storage	.12%	.12%	.11%
231	Type B - Unoccupied site for mobile home, trailer, or tent	.41%	.46%	.48%
232	Type B - Permit granted, construction not started	.06%	.05%	.04%
233	Type B - Other	.29%	.32%	.23%
	<b>Total type C</b>	<b>1.28%</b>	<b>1.00%</b>	<b>1.02%</b>
240	Type C - Demolished	.31%	.30%	.28%
241	Type C - House or trailer moved	.12%	.13%	.11%
242	Type C - Outside segment	.02%	.01%	.01%
243	Type C - Converted to permanent a business or storage	.13%	.11%	.12%

Outcome code	Description	Percent of cases		
		2011	2012	2013
244	Type C - Merged	.08%	.05%	.06%
245	Type C - Condemned	.02%	.02%	.02%
246	Type C - Built after April 1, 2000	.03%	.02%	.02%
247	Type C - Unused line of listing sheet	.12%	.06%	.09%
248	Type C - Other	.37%	.25%	.28%
256	Type C - Removed during subsampling	.00%	.00%	.00%
257	Type C - Unit already had a chance of selection	.08%	.05%	.03%

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2011, 2012, and 2013.

In addition to noninterviews at the sample address or household level, an eligible household member can be classified as a Type Z noninterview if the interviewer is unable to interview an eligible household member and the household member is not the household respondent. Unlike Type A, B, and C noninterviews, a Type Z noninterview relates to a household member, not to the entire sample household. The monthly average of eligible NCVS persons classified as Type Z noninterviews was 12.22% during the 2011 collection year, 13.20% in 2012, and 12.41% in 2013.

#### *Exclusions due to dangerous areas*

Field representatives play an important role in gathering NCVS information from sample persons or addresses. However, the safety of the field representatives is equally important. For this reason, the Census Bureau maintains a Dangerous Address Database (DAD) to keep track of addresses that may be a potential personal safety risk. Addresses are entered into the DAD based on information from the field staff or other sources. Each flagged address is assigned a status of “Interview with Caution” or “Cease Interview.” Once a case has been flagged, it maintains its DAD status throughout the duration of the survey sample or until the address is unflagged.

Addresses flagged as “Interview with Caution” could be potentially dangerous. However, interviewing may still be attempted. Examples of this status include a sample address flagged as dangerous due to circumstances involving a nearby nonsample address or because of a known personal safety risk within the building where the sample address is located.

Addresses flagged as “Cease Interview” are a serious personal safety risk and should not be attempted for interviewing. Depending on the situation, field representatives are instructed only to verify the occupants are still living in the unit through the post office, neighbor, or other usual sources, and/or confirm the dangerous situation still exists, and then transmit the case back to the regional office as a noninterview.

As of February 2012, approximately 37 NCVS addresses nationwide were classified as “Interview with Caution” and 26 NCVS addresses nationwide were classified as “Cease Interview.” This information is subject to change on a monthly basis.

## Response rates

### *Household response rates*

The overall household response rates were 90.2% in 2011, 86.7% in 2012, and 84.5% in 2013.

### *Person response rates by selected demographic characteristics*

The overall person response rates were 87.9% in 2011, 86.8% in 2012, and 87.6% in 2013.

Tables 3.6, 3.7, and 3.8 present response rates by demographic characteristics including age, race and ethnicity, and sex.

<b>Age</b>	<b>2011</b>	<b>2012</b>	<b>2013</b>
12–17	72.83	72.78	70.89
18–24	77.40	76.25	78.61
25–34	88.25	86.73	87.39
35–49	89.20	88.13	88.99
50–64	91.08	89.66	90.46
65 or older	95.27	94.41	94.97

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2011, 2012, and 2013.

<b>Sex</b>	<b>2011</b>	<b>2012</b>	<b>2013</b>
Male	86.32	85.13	86.13
Female	89.30	88.36	88.95

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2011, 2012, and 2013.

<b>Race and ethnicity</b>	<b>2011</b>	<b>2012</b>	<b>2013</b>
White*	88.77	87.59	88.38
Black/African American*	85.95	85.47	87.37
Hispanic/Latino	86.21	85.79	85.49
American Indian/ Alaska Native*	89.03	89.78	88.04
Asian*	84.85	83.02	85.13
Native Hawaiian/ other Pacific Islander*	81.06	77.62	84.67
Two or more races*	85.88	83.24	82.76

\*Excludes persons of Hispanic or Latino origin.

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2011, 2012, and 2013.

### Attrition and time-in-sample (measure of retention over time)

#### *Response rates by time-in-sample*

Tables 3.9 and 3.10 show household and person level response rates for 2011, 2012, and 2013 by time in sample (TIS). TIS 1 denotes the first month in sample, TIS 2 the second month in sample, and so on.

<b>Time in sample</b>	<b>2011</b>	<b>2012</b>	<b>2013</b>
TIS 1	88.78	85.45	83.07
TIS 2	89.59	86.17	84.60
TIS 3	89.86	86.85	83.92
TIS 4	90.51	86.04	84.33
TIS 5	90.70	86.31	84.14
TIS 6	91.38	87.24	84.81
TIS 7	92.24	89.58	87.41

Note: Time-in-sample (TIS) refers to the interview number of a given unit. For example, a unit in its third interview is referred to as TIS 3, or the unit's third time-in-sample.

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2011, 2012, and 2013.

<b>Time in sample</b>	<b>2011</b>	<b>2012</b>	<b>2013</b>
TIS 1	89.37	87.37	88.72
TIS 2	87.83	87.07	87.38
TIS 3	87.19	87.18	87.21
TIS 4	86.61	86.03	86.89
TIS 5	87.49	85.94	87.15
TIS 6	87.45	86.89	87.78
TIS 7	88.37	87.38	88.21

Note: Time-in-sample (TIS) refers to the interview number of a given unit. For example, a unit in its third interview is referred to as TIS 3, or the unit's third time-in-sample.

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2011, 2012, and 2013.

Table 3.11 presents the detailed survey response status for each TIS wave. These estimates were generated using a longitudinal file, with an incoming household cohort starting in the first half of 2010.

<b>Status</b>	<b>Wave 1</b>	<b>Wave 2</b>	<b>Wave 3</b>	<b>Wave 4</b>	<b>Wave 5</b>	<b>Wave 6</b>	<b>Wave 7</b>
Type A							
Interviewed	5,418	5,461	5,439	5,333	5,252	5,140	5,296
Refused	253	266	312	387	459	516	448
No one home	186	148	130	147	148	173	104
Other Type A	88	89	89	105	108	138	104
Type A subtotal	5,954	5,964	5,970	5,972	5,967	5,967	5,952
Refusal rate	4.25%	4.46%	5.23%	6.48%	7.69%	8.65%	7.53%
Type B	1,248	1,189	1,146	1,120	1,108	1,087	1,084
Type C	319	40	37	24	16	21	17
Total	7,512	7,193	7,153	7,116	7,091	7,075	7,053
Not matched	N/A*	319	359	396	421	437	459

\*Not applicable.

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2010–2013.

Some households were interviewed in an early wave but were classified as a Type A nonresponse case in the subsequent wave. Table 3.12 shows the survey response status for each wave after the initial wave, among households that were also interviewed in the previous wave.

<b>Interviewed at wave (t)</b>	<b>Interview status at wave (t+1)</b>						
	<b>Type A</b>				<b>Type B</b>	<b>Type C</b>	<b>Total</b>
	<b>Interviewed</b>	<b>Refused</b>	<b>No one home</b>	<b>Other Type A</b>			
t=1	4,906	107	74	56	265	10	5,418
t=2	4,986	123	69	52	224	7	5,461
t=3	4,886	165	90	60	235	3	5,439
t=4	4,808	161	76	60	223	4	5,332
t=5	4,688	165	105	73	220	1	5,252
t=6	4,746	103	45	46	197	3	5,140

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2010.

## Chapter 4. Processing and editing

### Processing overview

NCVS processing is performed on two cycles: monthly and quarterly. Editing and coding is done on a monthly basis, before any other processing. The remainder of the monthly processing is then performed. This includes pre-editing, matching to the control unit file, industry and occupation coding, and reformatting. When 3 months of data for a quarter are available, the three consecutive monthly files are merged and processed through the quarterly cycle, including (in order): range checking, consistency editing, blanking editing, type of crime classification, period-to-period incident recounting, industry and occupation allocation, splitter/family structure recoding, weighting, recoding, and stripping type of crime (TOC) code 90 from the NCVS incident file. All operations are explained in further detail later in this section.

### Editing and coding

The purpose of the editing and coding operation is to improve the accuracy and quality of the data by locating and correcting interviewer errors. Editing is a manual process in which clerks at the Census Bureau's National Processing Center (NPC) review every case containing at least one crime incident report and any respondent race and ethnicity entry of "Other-specify." Identified errors are corrected, coded, or referred for further resolution. Special cases and any unusual situations identified by clerks are referred to Census Bureau headquarters staff.

After an interviewing month has closed out, a file is created that is comprised of all cases containing at least one incident report. This file is then loaded for coders at the NPC to begin the review process. Coders at the NPC conduct a two-phase editing and coding review process. The first stage is the initial review of the crime incidents collected. During this stage, NPC coders compare incident data to incident summaries and any pertinent information in case notes.

Particular attention is paid to questions used to classify crime incidents, such as location, presence, physical attack, attempted attack, threat, and theft. In addition, any question containing an "Other-specify" write-in entry is reviewed and, when possible, recoded to one of the pre-coded categories. Respondent race and ethnicity entries of "Other-specify" are also reviewed and, when possible, recoded to one of the pre-coded categories. The initial reviewer can correct or accept the data or add a referral code to send the case to headquarters staff for further review.

When all of the cases have gone through the initial review, incidents then go through a second review, the verification phase, during which edits and referral codes are verified and any new referral reasons can be added. Incidents can be referred for one or more of 38 referral reasons. In most instances, referrals occur because collected data does not match the summary or other available information. After the verification stage, cases that do not have a referral linked to any incidents are closed.

After the NPC has completed both the initial and verification stages, any case containing an incident with at least one referral reason remains open for headquarters staff to review. In a typical month, about 1,000 cases are on the file sent to the NPC, of which about 300 are referred to the headquarters staff for resolution. The headquarters staff reviews all incidents with a referral reason in order to resolve any discrepancies, using the summary and other available information to make decisions. After headquarters staff have completed their review, the editing and coding file for the month is closed out.

All of the incidents that occurred in the victim's place of residence are assigned a Geographical Identification Code Scheme (GICS) code in post-data collection processing. The components of the GICS code are the Federal Information Processing Standards (FIPS) state, county, Minor Civil Division (MCD) (for MCD states in New England), and place codes. For incidents that occurred in a place differing from the victim's place of residence but still within the United States, headquarters staff assign the GICS code.

Some of the codes used to classify occupations are collapsed and include multiple occupations. For example, during the coding process for industry and occupation (I&O), coders classify elementary school teachers and middle school teachers with the same occupation code. However, BJS requested the ability to distinguish between middle school and elementary school teachers because this level of detail is collected in the employment section of the questionnaire. To accommodate this request, headquarters staff use the respondent's employment information to recode middle school teachers.

### *Monthly processing, following editing and coding operations*

- During the pre-edit stage of computer processing, basic data integrity checks are done. The pre-edit identifies households with line number errors, excessive numbers of persons (i.e., more than 13 persons age 12 or older), excessive numbers of incidents (i.e., more than 18 incident reports), and out-of-range incidents (i.e., before or after the reference period). These situations are flagged for correction during the data reformat stage. The pre-edit also verifies and codes whether the household was interviewed or was a noninterview; ensures that a valid race and ethnicity code was entered for noninterview, or Type A, households; resolves discrepancies over the interview or noninterview status of household members; checks for blank screeners; ensures the household respondent was interviewed; and corrects or deletes invalid age entries, verifies the counts of household members age 12 or older, and verifies incident reports. In addition, the pre-edit identifies households in the incoming rotation group.
- When the pre-edit is complete, each NCVS record is matched to the unit control file (UCF) to obtain more detailed geographic and demographic information. The entire UCF record is copied to the NCVS internal file. Most of this information is suppressed for disclosure reasons during the creation of the public use file, but these data are available on the restricted-use files archived at research data centers (RDC).
- Each month, all incidents that occurred while the respondent was working or on duty are extracted from the total universe of reported incidents and provided to the NPC coding unit for industry and occupation (I&O) coding, using the 4-digit North American Industry Classification System (NAICS) industry and 4-digit Standard Occupational Classification (SOC) occupation codes.
- The reformat, also known as the adjust program, is completed monthly, after performing the pre-edit and the match to the UCF. The adjust program applies corrective action to the edit failures identified during the pre-edit process. During the reformat process, machine corrections are made to the data based on errors flagged during the NCVS computer pre-edit. Machine corrections are made monthly and sometimes involve classifying the household as a Type A noninterview.

### *Quarterly data processing activities*

- A range check edit is performed on a quarterly basis to ensure a valid entry exists for each survey (i.e., questionnaire) item requiring a numeric entry. During the range check edit, a valid range is assigned for each data item, including acceptable codes for “blank.” If a data item does not contain an entry or if the entry is other than a valid code specified for that data item, a not available (NA) response code is entered for that item. For single entry items, an “8” is entered in the right most position of the data field, preceded by the appropriate number of “9s.” For multiple entry items, if one or more out-of-range entries are specified, an “8” is entered in the last position of the field. All data items are checked and resolved at this point in the operation.
- A consistency edit examines the responses to individual items. It determines if the responses are consistent with the other data on the questionnaire and follow a logical, reasonable response pattern. If inconsistent answers are present, they are changed or deleted based on other answers. For example, if no adequate entry is specified for sex, then sex is allocated to “male” based on relationship codes of husband, son, father, or brother. Answers that pass the consistency test remain unchanged but may be blanked or changed to a not available response code in later phases of processing. In addition, selected demographic data (e.g., age, sex, ace) are imputed for noninterviewed persons in interviewed households (i.e., Type Z person noninterview). This is also the stage at which allocation/imputation occurs using a hot-deck allocation.
- A blanking edit program checks to make sure each case follows the questionnaire skip pattern by stepping sequentially through the items. It deletes entries that should not have been filled, though in certain situations interviewers can bypass error messages about skip pattern violations that appear during the interview. Out-of-universe codes for items that should be blank are assigned and consist of a “9” in each position in the field for that item. For example, if a data item has a field length of five positions and the data item is out-of-universe, then the data item will contain a value of “99999.”
- Each criminal incident is assigned a type of crime (TOC) code that depends on the entries in the incident report. Each criminal incident is counted only once and is classified by the most serious act that took place during the incident, ranked in accordance with the seriousness classification system used by the FBI. The seriousness of crimes against persons is, in descending order, rape, sexual assault, robbery, aggravated assault, simple assault, and purse snatching/pocket picking. If a person is both robbed and assaulted, the event is classified as a robbery. If the victim suffers physical harm, the crime is categorized as robbery with injury. Personal crimes of contact take precedence over household offenses. Among the latter, household burglary is the considered most serious while personal theft is considered the least serious. All information about co-occurring incidents is retained, allowing each incident to be examined by reviewers. Incidents that cannot be classified according to the crime classification algorithm (e.g., arson, con-games, and kidnapping) are deleted from the file. (See page 5 for a detailed list of crimes not covered in the NCVS.)
- A period-to-period recount is conducted to determine which reported incidents (i.e., household and personal crimes) occurred in the month of the interview. Crimes that occurred during the month of the interview technically fall outside of the 6-month reference period and are therefore counted during the next interview period. These incidents are identified, placed in a hold file, and subsequently matched to data received

for the next reporting period before processing—6 months later. Personal incidents are added to the quarterly processing file, if the household and personal characteristics for the hold file record match the household and personal characteristics for the current file record. If not, then the incident is excluded from further processing. Household crimes from the hold file are added to the current quarterly processing file, if household characteristics match a household on the current month file. If not, then these incidents are also excluded from subsequent processing.

- Allocation of I&O codes was initiated, beginning with the July 2001 data, to identify incident records in which the respondent was working or on duty and the NPC could not assign an industry and/or occupation code. This process is performed quarterly after the period-to-period recount to ensure that all incident records, current and hold files, are identified for I&O coding. This program was modeled after the Current Population Survey (CPS) I&O allocation program. The assignment of I&O codes was based on the 1990 Standard Industrial Classification (SIC) and SOC coding system through December 2002. After December 2002, the process was migrated to the NAICS/SOC Coding System.
- The “family structure” recode provides information about the sex and marital status of the household respondent, as well as the relationship of other household members to the household respondent.
- Weighting is the process of adjusting the sample counts to correct differences between the sample and population totals. Weighting is done through a program that calculates weights for every interviewed household, each interviewed person, each victimization, and each reported incident. See Chapter 5 for detail on weighting adjustments.
- Beginning in January 2003, to comply with the Office of Management and Budget’s (OMB) 1997 directive on race and ethnicity, respondents could indicate multiple responses to the race and ethnicity item. At the request of the Census Bureau Disclosure Review Board (DRB), any reported race and ethnicity combinations must be represented in the total population by at least 10,000 persons. To meet this requirement, a race and ethnicity recode is done on the multiple-entry race and ethnicity item. For disclosure reasons, geographic variables are recoded to ensure that no data are published for geographic areas with populations less than 100,000 residents.
- All TOC 90, unwanted sexual contact crimes, are stripped from the NCVS incident file. Incidents of this type are not classified into any crime category and are not considered crimes for the purposes of the NCVS.

### *Imputation*

Imputation is used to assign values for missing items (i.e., characteristics) of proxy and self-interviewed sample persons, and to replace reported values that fail consistency edits. Imputation is also used to assign data to Type Z noninterviewed (i.e., direct or indirect refusal to interview) sample persons in interviewed sample households. A Type Z noninterviewed sample person is a person in an interviewed household for whom no information was collected from either the person or a proxy respondent.

The Census Bureau’s traditional, sequential hot-deck procedure is used to impute missing or rejected values for selected characteristics of interviewed persons. The variables (i.e., characteristics) used to define imputation matrices vary widely from survey to survey, depending on the item being imputed, and for the NCVS include age, sex, and race. For each missing value,

the procedure assigns a value reported for a person with similar characteristics, also known as a donor record. For each item subject to imputation, both the original and allocated values are kept on the data file.

Imputation is also used to assign all of the values for a Type Z noninterviewed sample person in an interviewed sample household. Complete records are imputed for Type Z nonrespondents using interviewed cases. However, crime data are not imputed. In addition, variables on the incident report may be allocated values based on consistency edits with other variables. Table 4.1 shows NCVS imputation rates for age, sex, and race.

<b>Table 4.1. NCVS imputation rates, by age, sex, and race, 2011, 2012, and 2013</b>						
	<b>2011</b>		<b>2012</b>		<b>2013</b>	
	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>	<b>Number</b>	<b>Percent</b>
Age	6,995	4.29%	6,723	3.58%	6,702	3.67%
Sex	1,369	0.84%	980	0.52%	1,019	0.56%
Race	72	0.04%	36	0.02%	48	0.03%

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2011, 2012, and 2013.

Data on household income in the NCVS is characterized by high item-nonresponse. Historically, data on income has not been imputed for the survey because Census has focused imputation efforts on variables used in weighting survey estimates.

## Chapter 5. Weighting

Household, person, and victimization data from the NCVS sample are adjusted to give quarterly and annual estimates of crime experienced by the U.S. household population age 12 or older.<sup>2</sup> Adjustments are first made to account for noninterviews. Additional adjustment factors then reduce the variance of the estimate by correcting for differences between the distribution of the sample by age, sex, and race and the distribution of the population by these characteristics. This also reduces bias due to undercoverage of various portions of the population, when up-to-date noninstitutional population projections from the Population Division of the Census Bureau are used for ratio adjustments.

This section discusses the various components of NCVS weights, which are summarized in table 5.1.

**Table 5.1. Components of NCVS weights**

Components of the NCVS weights	Household-level estimates			Person-level estimates		
	Household	Victimization	Incident	Person	Victimization	Incident
Base weight	×	×	×	×	×	×
Special weighting adjustments	×	×	×	×	×	×
Household noninterview factor	×	×	×	×	×	×
Within-household noninterview factor				×	×	×
First-stage ratio	×	×	×	×	×	×
Second-stage ratio	×	×	×	×	×	×
Bounding adjustment		×	×		×	×
Multiple victim adjustment						×

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2013.

<sup>2</sup>Throughout the remainder of this document, the terms "household" and "housing unit" are used interchangeably.

## Base weights

The original NCVS base weights at the beginning of the 2000 design were designed to be self-weighting (Kish, 1965). Self-weighting refers to a sample design in which the overall weights for all of the units in the sample are equal. With a two-stage sample design, the second-stage probabilities of selection are adjusted for the unequal first-stage probabilities of selection so that the product is constant. Different sample reductions and expansions during the course of the 2000 design have altered the self-weighting design.

## Special weighting adjustments

Some units are subsampled in the field because the observed size is much larger than expected. During the estimation procedure, housing units in these units must receive special weighting factors (i.e., weighting control factors) to account for the change in the probability of selection. For example, an area sample unit expected to have four housing units but found at the time of interview to contain 36 housing units, could be subsampled at the rate of 1 in 3 to reduce the interviewer's workload. Each of the 12 designated housing units in this case would be given a special weighting factor of 3. To limit the effect of this adjustment on the variance of sample estimates, these special weighting factors are limited to a maximum value of 4.

At this stage of the NCVS estimation process, the special weighting factors are multiplied by the base weights. The resulting weights are then used to produce "unbiased" estimates. Although this estimate is commonly called unbiased, it includes some negligible bias because the size of the special weighting factor is limited to 4. The purpose of this limitation is to achieve a compromise between a reduction in the bias and an increase in the variance.

## Adjustment for household nonresponse

Nonresponse arises when households selected for inclusion in a survey fail to provide all or some of the data that were to be collected. This failure to obtain complete results from all of the selected units can arise from several different sources, depending on the survey situation. Nonresponse is classified into two major types: item nonresponse and complete (or unit) nonresponse. Item nonresponse occurs when a cooperating household fails or refuses to provide some specific items of information.

Unit nonresponse refers to the failure to collect any survey data from an occupied sample household, which was about 16% in 2013. For example, data may not be obtained from an eligible household in the survey because of impassable roads, a respondent's absence or refusal to participate in the interview, or unavailability of the respondent for other reasons. Similarly, the within-household nonresponse adjustment addresses unit nonresponse of persons and not item nonresponse. In the NCVS estimation process, the weights for all of the interviewed households are adjusted to account for occupied sample households for which no information was obtained due to unit nonresponse.

Weighted counts of interviewed and noninterviewed households are tabulated separately for each noninterview adjustment cell. The weight for this purpose is the base weight multiplied by any special weighting factor. The household noninterview factor (HHNAF) is computed as—

$$HHNAF = \frac{A+B}{A}$$

where

$A$  = the weighted count of interviewed households

$B$  = the weighted count of noninterviewed households

Both  $A$  and  $B$  are weighted by the base weight times the special weighting factor. Cells are collapsed when the nonresponse factor is greater than or equal to 2.0. If collapsing is necessary, the weighted counts are combined, and a common adjustment factor is computed and applied to weights for interviewed persons within the collapsed cells.

To reduce estimate bias, the noninterview adjustment is performed within cells that are formed using the following variables:

- type of living quarters (i.e., a housing unit or not a housing unit)
- CBSA/Metropolitan Statistical Area (MSA) status (part of a principal city within a CBSA/MSA, not part of a principal city but still within a CBSA/MSA, or outside of a CBSA/MSA)
- urban status
- race of the reference person or head of household (white only, all other races and combinations of races).

These variables are cross-classified to make cells for the household nonresponse adjustment. Overall, 18 cells are specified because some cells are collapsed.

### **Adjustment for person nonresponse**

The within household noninterview factor (WHNAF) is computed as—

$$WHNAF = \frac{C+D}{C}$$

where

$C$  = the weighted count of interviewed persons

$D$  = the weighted count of noninterviewed persons

Both  $C$  and  $D$  are weighted by the base weight times the special weighting factor. Cells are collapsed when the nonresponse adjustment factor is greater than or equal to 2.0, or if the cells have less than 30 interviewed persons. If collapsing is necessary, the weighted counts are combined, and a common adjustment factor is computed and applied to weights for interviewed persons within the collapsed cells.

To reduce estimate bias, the noninterview adjustment is performed within cells that are formed using the following variables: region, age, sex, and race. These variables are cross-classified in different ways, depending on household relationship. Tables 5.2 and 5.3 summarize the cells and provide actual noninterview factors from one region in one quarter of 2013. To avoid potential disclosure, the Census Bureau will not identify the region and quarter used in these tables. Although not indicated in tables 5.2 and 5.3, the second row identifies the age of the person.

**Table 5.2. NCVS nonresponse cells and adjustment factors in 2013 for reference person and spouse**

Sex	Reference person and spouse							
	Black only				Other race			
	12–34	35–49	50–64	65 or older	12–34	35–49	50–64	65 or older
Male	1.1148	1.0822	1.0588	1.0182	1.0691	1.1046	1.1113	1.0630
Female	1.0701	1.0660	1.0862	1.0145	1.0536	1.0920	1.0723	1.0397

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2013.

**Table 5.3. NCVS nonresponse cells and adjustment factors in 2013 for all other persons - not reference person and spouse**

Sex	All other persons—not reference person and spouse									
	Black only					Other race				
	12–17	18–29	30–39	40–49	50 or older	12–17	18–29	30–39	40–49	50 or older
Male	1.5789	1.4600	1.4600	1.2727	1.2727	1.4638	1.4604	1.3750	1.3651	1.2600
Female	1.5128	1.3265	1.3265	1.1250	1.1250	1.4641	1.3928	1.3423	1.2727	1.1469

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2013.

### Ratio adjustment factors

Distributions of the demographic characteristics derived from the NCVS sample in any month will be somewhat different from the true distributions, even for such basic characteristics as age, sex, race and Hispanic or Latino origin. These particular population characteristics are closely correlated with victimization status and other characteristics estimated from the sample.

Therefore, the variance of sample estimates based on these characteristics can be reduced when, by the use of appropriate weighting adjustments, the sample population distribution is brought as closely into agreement as possible with the known distribution of the entire population, with respect to these characteristics. This is accomplished by means of ratio adjustments.

Two ratio adjustments are used in the NCVS estimation process: the first-stage ratio adjustment and the second-stage ratio adjustment. In the first-stage ratio adjustment, weights are adjusted so that the distribution of the single-race black/African American population and the population that is not single-race black/African American (based on the census) in a state's sample PSUs correspond to the same population groups' census distribution in all PSUs in the state. In the second-stage ratio adjustment, weights are adjusted so that aggregated NCVS sample estimates match independent estimates of population in various age/sex/race and age/sex/ethnicity cells at the national level.

### First-stage ratio adjustment

The purpose of the first-stage ratio adjustment is to reduce the variance due to the first-stage sample design, which is sometimes referred to as the between-PSU variance.

Several factors are considered when determining which information to use in applying the first-stage adjustment. The information must be available for each PSU, correlated with as many of the statistics of importance published from the NCVS as possible, and reasonably stable over time so that the accuracy gained from the ratio adjustment procedure does not deteriorate.

This adjustment is not necessary for self-representing (SR) PSUs because they represent only themselves.

### Computing first-stage ratio adjustment factors

The first-stage adjustment factors are based on Census 2000 data and are applied only to sample data for the non self-representing (NSR) PSUs. The first-stage adjustment factors (FSF) are calculated as—

$$FSF = \frac{\sum_{i=1}^{N_1} N_{personsi}}{\sum_{i=1}^{n_1} \frac{N_{personsi}}{\pi_i}}$$

where

$i$  = the index on the PSUs in the sum

$N_{personsi}$  = the known civilian noninstitutional population for PSU  $i$

$\pi_i$  = the probability of selection for PSU  $i$

$N_1$  = the number of first-stage PSUs in the first-stage universe

$n_1$  = the number of first-stage PSUs in the first-stage sample

To reduce estimate bias, the first-stage adjustment is performed within cells that are formed using the following variables:

- region
- principal city of a CBSA/MSA, balance of a CBSA/MSA, or not in a CBSA/MSA
- race.

### Second-stage ratio adjustment

The second-stage ratio adjustment decreases the error in the great majority of sample estimates. The procedure reduces some coverage error bias by calibrating sample estimates to independent population controls, which are updated each month. The procedure adjusts the weights to control the sample estimates for a number of geographic and demographic subgroups of the population to ensure that these sample-based estimates of the population match independent population controls in each of these categories.

The second-stage ratio adjustment (SSF) is calculated as—

$$SSF = \frac{N_{persons}}{\hat{N}_{persons}}$$

where

$\hat{N}_{persons}$  = the estimated Census 2000 civilian noninstitutional population from the survey

$N_{persons}$  = the known Census 2000 civilian noninstitutional population

The weighted count is the weighted sum of all persons where the weight is defined by the product (base weight) × (special weighting factor) × (household nonresponse adjustment factor) × (within household nonresponse adjustment factor) × (first-stage ratio adjustment factor). To reduce estimate bias, the second-stage adjustment is performed within cells that are formed using the following variables:

- Age
- Sex
- Race

Cells are collapsed when the second-stage ratio adjustment factor is less than 0.5, greater than or equal to 2.0, or if the cells have less than 30 interviewed persons. If collapsing is necessary, the weighted counts are combined and a common adjustment factor is computed and applied to weights for interviewed persons within the collapsed cells.

Table 5.4 presents the average coverage ratios for 2013. For example, the value of 0.85 for female says that the weights of persons in this cell were increased by a factor of  $1 / 0.85 = 1.18$  to account for the 15% of population that is female and is not represented in the estimates. For the coverage ratios for each month in 2013, see Appendix B.

**Table 5.4: Average Coverage Ratios in 2013**

Demographic characteristic	Coverage ratio
<b>Age</b>	
12–17	0.79
18–24	0.68
25–34	0.80
35–49	0.85
50–64	0.89
65 or older	0.93
<b>Sex</b>	
Male	0.83
Female	0.85
<b>Race/Hispanic origin</b>	
White	0.81
Black/African American	0.87
Hispanic/Latino	0.77
American Indian/Alaska Native	0.63
Asian/Native Hawaiian/Other Pacific Islander	0.82
Two or more races	0.65

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2013.

## Sources of independent controls

The independent population controls of  $N_{persons}$  are prepared by projecting forward the population figures derived from Census 2000, using information from a variety of other sources that account for births, deaths, and net migration. The civilian noninstitutionalized population is calculated by subtracting estimated numbers of resident armed forces personnel and institutionalized persons from the resident population. Prepared in this manner, the controls are themselves estimates. However, controls are derived independently of the NCVS and provide useful information for adjusting sample estimates.

## Bounding adjustment

Telescoping occurs when respondents report events that fall outside of the period of interest. (See Neter & Waksberg (1964) for more information on telescoping.) Telescoping causes overreporting and often happens in surveys when respondents are asked to recall all events within a given period. NCVS asks respondents to recall all incidents that occurred during the previous 6 months. Prior to 2007, the first NCVS interview was a bounding interview and was not used in estimates, to avoid potential telescoping bias. Since 2007, the first of the seven NCVS interviews has been used in estimates, in conjunction with a bounding adjustment for the first interview, to avoid telescoping bias.

In addition, NCVS has a known time-in-sample effect, wherein respondents in earlier interviews report more incidents than respondents in later interviews. Therefore, adding the first interview without an adjustment would result in increased incident rates, compared to incident rates of prior survey years. (See Biderman & Cantor (1984) for more information on time-in-sample bias.) To make data from all seven interviews comparable to data from the six interviews, an adjustment is applied to the weights of units in the first interview. Because respondents in the first interview report more incidents than in subsequent interviews, the adjustment reduces incident reporting in the first interview. The adjustment is calculated separately for household incidents and person incidents.

The adjustment makes estimated incident rates since 2007 comparable to estimated incident rates prior to 2007 by reducing the incident rate of the first interview to be consistent with the mean of the second through seventh interviews. This is completed by applying a factor to the weight of all of the units in the first interview. The bounding adjustment factor for incidents of person types ( $BAF_{pi}$ ) is defined as—

$$BAF_{\substack{person \\ incidents}} = \frac{\frac{1}{6} [\hat{R}_{persons}^{(2)} + \hat{R}_{persons}^{(3)} + \hat{R}_{persons}^{(4)} + \hat{R}_{persons}^{(5)} + \hat{R}_{persons}^{(6)} + \hat{R}_{persons}^{(7)}]}{\hat{R}_{persons}^{(1)}}$$

The bounding adjustment factor for incidents of household types ( $BAF_{HHi}$ ) is defined as—

$$BAF_{\substack{HH \\ incidents}} = \frac{\frac{1}{6} [\hat{R}_{HH}^{(2)} + \hat{R}_{HH}^{(3)} + \hat{R}_{HH}^{(4)} + \hat{R}_{HH}^{(5)} + \hat{R}_{HH}^{(6)} + \hat{R}_{HH}^{(7)}]}{\hat{R}_{HH}^{(1)}}$$

where  $\hat{R}_{persons}^{(2)}$  is the person incident rate and  $\hat{R}_{HH}^{(i)}$  is the household incident rate for the  $i^{th}$  interview.

The bounding adjustment is applied to both incident and victimization weights. Overall, two bounding adjustments are used: one for persons and one for households. The personal crime bounding adjustment should be used when estimating all types of personal crimes. The household bounding adjustment should be used when estimating all types of household crimes.

### **Adjustment for multiple victims**

If every victimization had one victim, the incident weight would be the same as the victimization weight. Because incidents often have more than one victim, to get the incident weight, the victimization weight is multiplied by the following multiple victim factor (MVF):

$$MVF = \frac{1}{\text{Number of victims in the incident}}$$

The number of victims in an incident can be greater than the number of persons in the household. Note that for household crimes, the incident weight and the victimization weight will always be same.

### **Adjustment for series crimes**

For series crimes, the weight counts series incidents as the actual number of incidents reported by the victim, up to a maximum of 10 incidents. In this case, the incident weight is the product of the victimization weight and the number of crimes in the series, with the number of crimes in the series capped at 10. Including series victimizations in national rates results in rather large increases in the level of violent victimization; however, trends in violence are generally similar regardless of whether series victimizations are included.

### **Household weight**

One way to define the household weight is to use the adjusted weight of one of the persons in the household. NCVS does this by defining the household weight as the weight of the principal person. In husband–wife households, the principal person is the wife. In other household types, the principal person is the reference person. (See Appendix A: Terms and definitions, for the definition of “reference person.”) If the principal person is a nonrespondent, then the weight that would have been assigned if the person had been interviewed is used, excluding the person-level, non-interview adjustment. Because the personal characteristics of the principal person (i.e., age, sex, and race) is known, the weight the person would have received if he or she had been interviewed is also known.

### **Household incident weight**

The victimization weight for the household is the same as the household weight because only one household is used.

### **Design weights for NCVS supplements**

Design weights specific to the supplements fielded by BJS are produced so that the same types of totals and rates as those described in the previous section may be estimated for the supplements. Most supplements are at the person level and are not intended to produce household estimates. If everyone in the NCVS household that responded would also respond to the NCVS supplement, the person weight for the supplement would be the same as the person weight for NCVS. Because persons within households do not always respond, non-interviews are accounted for

with a person-level nonresponse factor, which is a person nonresponse adjustment for the supplement (different from the NCVS person nonresponse adjustment).

A unique second-stage ratio adjustment is not applied to the weights for supplements since a “known” total for the specialized universe of interest is not easily available.

### Weighting summaries

This section describes how weighting progressed from base weight (i.e., inverse of the probability of selection) to the final household and person-weights, which were post-stratified to match the control totals obtained from current forward projections from Census 2000.

Table 5.5 shows household base weight by minimum, first quartile, mean, median, third quartile, maximum, and total weight for each of the selected cells, which were chosen by crossing CBSA status and location of residence (i.e., urban, suburban, and rural). Ideally, in a self-weighting sample design, all weights are equal and the total weight represents the population total used as the measure of size in the probability sample. However, the design was altered due to sample reduction/expansion and other adjustments typical of multi-stage design. Therefore, small variations in weights are expected and observed (e.g., the observed first and third quartiles are similar). Table 5.6 shows the distribution of household weights after adjusting for nonresponse and subsampling. Table 5.7 shows the distribution of final household weights by CBSA status and location of residence.

**Table 5.5. NCVS household base weight, by CBSA status and location of residence, 2013**

		Base weight						
CBSA status	Location of residence	Minimum	First quartile	Mean	Median	Third quartile	Maximum	Total
Central city of a CBSA	Urban	1,008	1,008	1,008	1,008	1,008	1,008	34,069,759
Central city of a CBSA	Rural	1,008	1,008	1,008	1,008	1,008	1,008	2,060,220
Balance of CBSA	Suburban	1,008	1,008	1,008	1,008	1,008	1,008	48,782,731
Balance of CBSA	Rural	1,008	1,008	1,008	1,008	1,008	1,008	6,951,098
Not in a CBSA	Rural	1,008	1,008	1,008	1,008	1,008	1,008	16,419,234

Note: A Core-Based Statistical Area (CBSA) is a geographic area consisting of the county or counties or equivalent entities associated with at least one core (urbanized area or urban cluster) of at least 10,000 population, plus adjacent counties having a high degree of social and economic integration with the core as measured through commuting ties with the counties associated with the core.

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2013.

**Table 5.6. NCVS household weight adjusted for nonresponse and subsampling, by CBSA status and location of residence, 2013**

CBSA status	Location of residence	Household weight adjusted for nonresponse and subsampling						
		Minimum	First quartile	Mean	Median	Third quartile	Maximum	Total
Central city of a CBSA	Urban	1,008	1,145	1,203	1,198	1,245	3,025	34,051,245
Central city of a CBSA	Rural	1,008	1,130	1,223	1,168	1,293	2,169	2,119,071
Balance of CBSA	Suburban	1,008	1,140	1,211	1,194	1,254	6,319	49,093,026
Balance of CBSA	Rural	1,008	1,118	1,186	1,171	1,228	1,815	7,055,537
Not in a CBSA	Rural	1,008	1,086	1,166	1,144	1,211	3,135	16,439,133

Note: A Core-Based Statistical Area (CBSA) is a geographic area consisting of the county or counties or equivalent entities associated with at least one core (urbanized area or urban cluster) of at least 10,000 population, plus adjacent counties having a high degree of social and economic integration with the core as measured through commuting ties with the counties associated with the core.

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2013.

**Table 5.7. NCVS final household weight, by CBSA status and location of residence, 2013**

CBSA status	Location of residence	Final household weight						
		Minimum	First quartile	Mean	Median	Third quartile	Maximum	Total
Central city of a CBSA	Urban	844	1,306	1,447	1,398	1,528	4,159	40,957,187
Central city of a CBSA	Rural	902	1,305	1,465	1,404	1,558	3,374	2,537,860
Balance of CBSA	Suburban	719	1,289	1,415	1,371	1,482	8,561	57,384,879
Balance of CBSA	Rural	719	1,284	1,407	1,378	1,490	3,327	8,373,444
Not in a CBSA	Rural	867	1,153	1,302	1,243	1,393	3,868	18,368,947

Note: A Core-Based Statistical Area (CBSA) is a geographic area consisting of the county or counties or equivalent entities associated with at least one core (urbanized area or urban cluster) of at least 10,000 population, plus adjacent counties having a high degree of social and economic integration with the core as measured through commuting ties with the counties associated with the core.

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2013.

Similar to household weights, person weights also progressed as expected from a more stable base weight to slightly less stable final weights, as seen in tables 5.8, 5.9, and 5.10. Race (i.e., black/African American and non-black/African American) and sex were chosen to create cells to illustrate the weight progression.

**Table 5.8. NCVS person-level base weight, by race and sex, 2013**

		Base weight						
Race	Sex	Minimum	First quartile	Mean	Median	Third quartile	Maximum	Total
Black/African American	Male	1,008	1,008	1,008	1,008	1,008	1,008	9,041,570
Black/African American	Female	1,008	1,008	1,008	1,008	1,008	1,008	11,418,436
Non-black/African American	Male	1,008	1,008	1,008	1,008	1,008	1,008	79,190,886
Non-black/African American	Female	1,008	1,008	1,008	1,008	1,008	1,008	84,587,995

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2013.

**Table 5.9. NCVS person-level weight adjusted for nonresponse and subsampling, by race and sex, 2013**

		Weight adjusted for nonresponse and subsampling						
Race	Sex	Minimum	First quartile	Mean	Median	Third quartile	Maximum	Total
Black/African American	Male	1,027	1,087	1,184	1,112	1,273	4,424	9,054,142
Black/African American	Female	1,008	1,052	1,136	1,080	1,128	5,597	11,450,908
Non-black/African American	Male	1,049	1,083	1,175	1,114	1,159	6,222	79,538,457
Non-black/American	Female	1,030	1,067	1,139	1,090	1,129	5,990	84,993,587

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2013.

**Table 5.10. NCVS person-level final weight, by race and sex, 2013**

		Final person weight						
Race	Sex	Minimum	First quartile	Mean	Median	Third quartile	Maximum	Total
Black/African American	Male	817	1,504	1,993	1,868	2,362	7,643	15,247,925
Black/African American	Female	767	1,421	1,781	1,671	2,032	11,214	17,945,739
Non-black/African American	Male	976	1,424	1,676	1,588	1,820	10,642	113,486,470
Non-black/African American	Female	934	1,364	1,578	1,487	1,667	10,193	117,731,568

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2013.

## Chapter 6. Variance estimates

As with most large-scale household surveys, the NCVS uses a complex sample design involving stratification, multi-stage sampling, and unequal sampling rates. Weights are needed in the analysis to compensate for unequal sampling rates and nonresponse. Furthermore, most estimates from complex samples are nonlinear statistics. As a result, standard error estimates are often obtained using the first-order Taylor series approximations or replication methods, such as balanced repeated replication (BRR) or jackknife replication. This complex sample design needs to be considered when estimating the precision of survey estimates. Not accounting for these sample design features will lead to inaccurate point estimates and an underestimation of the precision. This section describes methods used for generating point estimates and variance estimates.

### Variance estimates with replication

Replication methods provide estimates of variance for a wide variety of designs using probability sampling, even when complex estimation procedures are used. This method requires the sample selection, data collection, and estimation procedures to be carried out (i.e., replicated) several times. Dispersing the resulting estimates can be used to measure the variance of the full sample.

### Two methods of variance replication

Two types of replicate variance estimation techniques are used in NCVS variance estimation: balanced repeated replication (BRR) and successive difference replication (SDR). Both of the techniques are embodied by the replicate factors produced by the NCVS replicate variance estimator.

### Replicate weights

The unbiased weights (baseweight  $\times$  special weighting factor) are multiplied by the replicate factors to produce unbiased replicate weights. These unbiased replicate weights are further adjusted through the noninterview adjustment, the first-stage ratio adjustment, and the second-stage ratio adjustments in the same manner as the full sample is weighted. By applying the other weighting adjustments to each replicate, the final replicate weights reflect the impact of the weighting adjustments on the variance.

Replicate weights are calculated for each of the four types of weights (i.e., household, person, incident, and victimization).

### Replicate factors for NSR strata

Replicate factors based on the BRR variance estimator are used for NSR strata (McCarthy, 1966). These replicate factors are used to measure the variance due to the selection of the first-stage sample. No PSUs were selected in SR strata. As a result, BRR replicate factors are not appropriate.

Because between-PSU variance cannot be estimated directly using BRR, it is instead calculated as the difference between the estimates of total variance and within-PSU variance. NSR strata are combined into pseudo strata within each state, and one NSR PSU from the pseudo stratum is randomly assigned to each panel of the replicate. Replicate factors of 1.5 or 0.5 adjust the weights for the NSR sample units. These factors are assigned based on a single row from a Hadamard matrix and are further adjusted to account for the unequal sizes of the original strata

within the pseudo-stratum. (See Wolter (1985) for more information on Hadamard matrices and BRR.)

These factors were further adjusted to account for the unequal sizes of the original strata within the pseudo stratum. All units in a pseudo stratum are assigned the same row number(s).

### Replicate factors for SR strata

The theoretical basis for the successive difference method was discussed by Wolter (1984) and extended by Fay and Train (1995) to produce the successive difference replication (SDR) method used for the NCVS. Because the variation of the SR PSUs comes entirely from selecting units within the PSU, the SDR method for SR PSUs is used, as defined by Fay and Train (1995). The following is a description of this method.

To apply SDR to the SR sample, the SR sample is sorted by PSU and then is sorted within PSU by the same order that was used to select the original *sys* sample. Each sample unit is then assigned two rows of the given Hadamard matrix. For example, the assignment for a Hadamard of order 160 would be rows (1,2) assigned to the first unit, rows (2,3) assigned to the second unit, ... rows (160,1) assigned to the 160<sup>th</sup> unit. The assignment is repeated in further cycles until the entire sample is assigned two rows.

For an SR sample, two rows of the Hadamard matrix are assigned to each pair of units, creating replicate factors,  $f_r$  for  $r = 1, \dots, R$  as—

$$f_{i,r} = 1 + 2^{-\frac{3}{2}} h_{i+1,r} - 2^{-\frac{3}{2}} h_{i+2,r}$$

where

$i$  = the index on the units of the sample

$r$  = the index on the set of replicates

$h_{i,r}$  = the number in the Hadamard matrix (+1 or -1) for the  $i^{\text{th}}$  unit in the systematic sample

$R$  = the number of total replicate samples or simply replicates

This formula yields replicate factors of approximately 1.7, 1.0, or 0.3.

### *Example 1: Successive difference replication*

The following simple example shows the SDR method. The sample in table 6.1 contains the weights of 5 units ( $n = 5$ ).

**Table 6.1. Sample weights**

Sample unit	Sample weight
Unit #1	15.00
Unit #2	23.00
Unit #3	19.00
Unit #4	16.00
Unit #5	21.00

The replicate factors are defined in the following  $4 \times 4$  Hadamard matrix:

$$\mathbf{H}_4 = \begin{bmatrix} +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 \end{bmatrix}$$

Two consecutive rows of  $\mathbf{H}_4$  are assigned to each sample unit, as denoted in table 6.2.

**Table 6.2. Assignment of rows in the Hadamard matrix**

Sample unit	Sample weight	Row I	Row II
Unit #1	15.00	2	3
Unit #2	23.00	3	4
Unit #3	19.00	4	2
Unit #4	16.00	2	4
Unit #5	21.00	4	3

Plugging these values into the replicate factor formula (1) generates—:

$$f_{11} = 1 + 2^{-\frac{3}{2}} h_{2,1} - 2^{-\frac{3}{2}} h_{3,1} = 1 + 2^{-\frac{3}{2}}(+1) - 2^{-\frac{3}{2}}(+1) = 1.0$$

$$f_{1,2} = 1 + 2^{-\frac{3}{2}} h_{2,2} - 2^{-\frac{3}{2}} h_{3,2} = 1 + 2^{-\frac{3}{2}}(-1) - 2^{-\frac{3}{2}}(+1) = 1 - \frac{1}{\sqrt{2}} \cong 0.3$$

$$f_{1,3} = 1 + 2^{\frac{3}{2}} h_{2,3} - 2^{\frac{3}{2}} h_{3,3} = 1 + 2^{\frac{3}{2}}(+1) - 2^{\frac{3}{2}}(-1) = 1 + \frac{1}{\sqrt{2}} \cong 1.7$$

Table 6.3 shows the calculated replicate factors for each replicate and unit in the sample.

**Table 6.3. Replicate factors**

Sample unit	Replicate factors			
	Replicate 1	Replicate 2	Replicate 3	Replicate 4
Unit #1	1.0	0.3	1.7	1.0
Unit #2	1.0	1.7	1.0	0.3
Unit #3	1.0	1.0	0.3	1.7
Unit #4	1.0	1.0	1.7	0.3
Unit #5	1.0	0.3	1.0	1.7

To calculate replicate weights, the full sample is multiplied by corresponding factors, as shown in table 6.4.

**Table 6.4. Assignment of rows in the Hadamard matrix**

Sample unit	Full sample weight	Replicate weights			
		Replicate 1	Replicate 2	Replicate 3	Replicate 4
Unit #1	15.0	4.5	25.5	15.0	15.0
Unit #2	23.0	39.1	23.0	6.9	23.0
Unit #3	19.0	19.0	5.7	32.3	19.0
Unit #4	16.0	16.0	27.2	4.8	16.0
Unit #5	21.0	6.3	21.0	35.7	21.0
Sum of weights	94.0	84.9	102.4	94.7	94.0

Other weighting adjustments for replicate weights

In example 1, adjustment ends at the point of adjusting the replicate base weights for the different replicates. The next step is to calculate the rest of the weighting adjustments for each set of replicate weights. The replicate weights also account for the effect on the variance of the other weighting factors. Recalculating the noninterview and second-stage ratio adjustments for each replicate ensures that the randomness injected or mitigated by the different weighting

adjustments is represented in each of the replicate estimates. (See Judkins (1990, p. 224) and Brick and Kalton (1996) for additional discussion of application of other weighting adjustments within replicate weighting.)

### Estimating generalized variance functions

NCVS uses a three-parameter model that is similar to the two-parameter model of the Current Population Survey (CPS). (See U.S. Census (2006) for more information on the motivation of the CPS two-parameter model.) Prior research showed that NCVS estimates had a better fit to the three-parameter model than the two-parameter model (Krenzke, 1995).

#### Generalization method

Let  $\hat{N}$  be an estimator of a total number of units (sum of the weights) within a domain interest and  $v(\hat{N})$  be its variance. The variance of  $v(\hat{N})$  is then modeled as a function of  $\hat{N}$  as

$$\frac{v(\hat{N})}{\hat{N}^2} = a + \frac{b}{\hat{N}} + \frac{c}{\sqrt{\hat{N}}} \quad (2)$$

where  $a$ ,  $b$ , and  $c$  are the parameters of the model.

For many domains of interest,  $\hat{N}$  is an estimator of a control total that used in the second-stage ratio adjustment. In these domains, as  $\hat{N}$  approaches  $N$ , the variance of  $\hat{N}$  approaches zero, since the second-stage ratio adjustment guarantees that these sample population estimates match independent population controls. The variance estimator assumes no variance on control totals, even though they are estimates. The generalized variance function (GVF) model satisfies this condition. The model and the GVF in general are only appropriate for estimating variances of totals like  $\hat{N}$  and are not appropriate for other statistics like  $\hat{Y}$  or rates. Variances of estimates based on totals of a characteristic of interest, rates or other statistics would likely fit a different functional form better.

The GVF parameters that accounts for a second-stage adjustment can be expressed as

$$\frac{v(\hat{N})}{\hat{N}^2} = b \left( \frac{1}{\hat{N}} - \frac{1}{N} \right) + c \left( \frac{1}{\sqrt{\hat{N}}} - \frac{1}{\sqrt{N}} \right) \quad (3)$$

(3) may then be derived from (2) by noting with the second-stage adjustment,  $v(\hat{N})|_{\hat{N}=N} = 0$ ,

therefore from (2),  $0 = a + \frac{b}{N} + \frac{c}{\sqrt{N}}$  or  $a = -\frac{b}{N} - \frac{c}{\sqrt{N}}$ . Substituting  $a$  into (2) gets (3).

The  $a$ ,  $b$  and  $c$  parameters are estimated by fitting a model to a group of related estimates and their estimated relative variances – the ratio of the variance and the square root of the estimate. The relative variances are calculated using the SDR method.

The model fitting technique is an iterative weighted least squares procedure, where the weight is the inverse of the square of the predicted relative variance. The use of these weights prevents items with large relative variances from unduly influencing the estimates of the  $a$ ,  $b$  and  $c$  parameters.

A year's worth of data is used in this model fitting process and each group of items should comprise at least 20 characteristics with their relative variances, although occasionally fewer characteristics are used.

Direct estimates of relative variances are required for estimates covering a wide range, so that observations are available to ensure a good fit of the model at high, low, and intermediate levels of the estimates. Using a model to estimate the relative variance of an estimate in this way introduces some error, since the model may substantially and erroneously modify some legitimately extreme values.

### *The utility of a three-parameter model*

Other surveys use a simpler model with only two parameters. For example, the CPS uses the two-parameter model defined as—

$$\frac{v(\hat{N})}{\hat{N}^2} = a + \frac{b}{\hat{N}}$$

Prior research indicated that the three-parameter model had a better fit than a two-parameter model for NCVS data. Mathematically, the two-parameter model defines a parabola, and the three-parameter model defines a leaning parabola, which has a steeper slope near zero.

### **Generalized variances for NCVS supplements**

Prior to 2008, the GVF was calculated using variances from a NCVS half-year sample, which used NCVS variables of interest and NCVS sample units. A NCVS half-year sample was used because supplements are usually collected within a half of a year. Also, the sample for supplements is often restricted on demographics and is not the full NCVS sample.

Since 2008, the methodology has calculated variances using the supplement sample and the associated supplement weights, coupled with NCVS variables of interest. The resultant variances are used to calculate the GVF.

The methods prior to and after 2008 assume that variances for supplement variables of interest are similar to variances for NCVS variables of interest. This should be true because most of the variation in supplement estimates come from the NCVS sample design.

### **Year-to-year correlations**

The following notation is used for year-to-year correlations:

$t, s$  = the indices of the time of the estimate (for NCVS, the indices  $t$  and  $s$  represent the year)

$r$  = the index on the replicates

$R$  = the number of replicates

$\hat{\theta}_t$  = the estimator of  $\theta$  at time  $t$

$\hat{\theta}_{t,r}$  = the replicate estimator of  $\theta$  at time  $t$  and replicate  $r$

The estimator for the year-to-year correlation for a general statistic  $\theta$  between time  $t$  and  $t - s$  is—

$$\hat{\rho}(\hat{\theta}_t, \hat{\theta}_{t-s}) = \frac{\text{cov}(\hat{\theta}_t, \hat{\theta}_{t-s})}{\sqrt{\hat{v}(\hat{\theta}_t)}\sqrt{\hat{v}(\hat{\theta}_{t-s})}}$$

which is defined in terms of the sample replicate covariances and variances that are functions of the replicate estimator of  $\theta_t$  and  $\theta_{t-s}$  at times  $t$  and  $t - s$ . That is—

$$\text{cov}(\hat{\theta}_t, \hat{\theta}_{t-s}) = \frac{4}{R} \sum_{r=1}^R (\hat{\theta}_{t,r} - \hat{\theta}_t)(\hat{\theta}_{t-s,r} - \hat{\theta}_{t-s})$$

and the variance of  $\hat{\theta}_t$  at time  $t$  is estimated with replication as—

$$\hat{v}(\hat{\theta}_t) = \frac{4}{R} \sum_{r=1}^R (\hat{\theta}_{t,r} - \hat{\theta}_t)^2$$

Assume that  $\hat{\theta}_t$  can be a nonlinear estimator, in that it is a nonlinear function of estimated totals. Appendices E and F contain the 1-year and 2-year correlations for 2013 and data from 2010 through 2013.

### Design effects

Because the NCVS departs from a simple random sample by using a complex sample design that includes stratification, clustering, and weighting, it is not as efficient as a simple random sample. Estimates are less precise due to these sample design features, which control cost. The complex design impacts the design effect (i.e., the ratio of the variance of the statistic from the complex design to the variance that would have been obtained had the sample been a simple random sample). Each estimate will have a different design effect depending on how the sample design affects the estimate. Clustering tends to increase the variance of survey estimates because the observations within a cluster are more homogenous than a random selection of observations. Clustered samples provide less information and are, therefore, not as precise.

Starting in 2011, the Census Bureau provided replicate weights, GVF parameters, and correlations, thereby enabling analysts to calculate variances and design effects.

## Appendix A: Terms and definitions

This section provides the definitions of several terms used with this technical document. It includes terms related to statistics, sample design, crime, and victimization. A more comprehensive list of terms and definitions specific to victimization can be found on the Bureau of Justice Statistics' (BJS) website <<http://www.bjs.gov/index.cfm?ty=tdtp&tid=9>>.

**2000 sample design.** The sample design implemented to select the samples used to calculate estimates from 2006 through 2015. The sample design is referred to as 2000 because it followed directly after the 2000 Census and used many of the data products from the 2000 Census to select an up-to-date sample.

**Area frame.** The sampling frame used in blocks where the unit frame has poor coverage or the permit frame is unavailable because the area does not have a Building Permit Office. Initially, it is a list of the blocks and the number of expected units within the block, according to the prior census. A sample of blocks is selected from the list, with probability proportional to size. Sampled blocks are sent for listing, from which the final sample of units is identified.

**Bias.** The formal definition of bias of an estimator  $\hat{\theta}$  of some statistic  $\theta$  is the expected value of the absolute value of the difference between the estimator and statistic and its expected value. That is,  $B(\hat{\theta}) = E|\hat{\theta} - \theta|$ . Informally, bias is a measurement of how close the estimator is to the value it is estimating.

**Balanced repeated replication (BRR).** A method of variance estimation often used with two-stage sample designs that select one or two PSUs per first-stage strata. This method is valuable because it can be applied to estimating the variance of linear and nonlinear estimates. Also, the intermediate replicate weights can be provided to data users, thereby enabling users to calculate estimate variances with simple expressions for the variance. The main ideas of replication are outlined by McCarthy (1966).

**Bounding.** A process to ensure that previously reported incidents are not reported again in the enumeration that follows. This is known as telescoping, or bringing incidents outside of the reference period into the period. Bounding provides a more accurate measure of criminal victimization within NCVS sample households.

**Calibration.** As described by Deville and Särndal (1992), calibration is a technique that can be used to reduce the variance of an estimator. Sometimes it can also have the effect of improving the coverage of the estimator. Calibration uses a set of known totals: either "imported totals" (Särndal & Lundström, 2005, p. 54) or a set of variables that are known for all units in the universe. Calibration finds weights close to the original design weights so that the estimated known totals with the new weights are the same as the known total.

**Central city/Balance/Urban/Rural.** A geographic identifier that indicates whether a sample block is in a central city of a 1999 MSA (i.e., the definition used for the 2000 Census), a balance of an (2000 population based) urbanized area, an urban cluster, or a rural cluster (i.e., outside the urbanized area or cluster).

**Coefficient of variation (CV).** The variance of an estimate divided by the square of the estimate. That is,  $v(\hat{\theta})/\hat{\theta}^2$ .

**Collection year.** The set of victimizations reported to the NCVS in interviews conducted during the same calendar year. This set may include victimizations that occurred in the previous calendar year because of the retrospective nature of the NCVS interview. BJS uses calendar year estimates in NCVS reports. See “Data year.”

**Core-Based Statistical Area (CBSA).** Metropolitan and micropolitan statistical areas (metro and micro areas) are geographic entities delineated by the Office of Management and Budget (OMB) for use by Federal statistical agencies in collecting, tabulating, and publishing Federal statistics. The term "Core Based Statistical Area" (CBSA) is a collective term for both metro and micro areas. A metro area contains a core urban area of 50,000 or more population, and a micro area contains an urban core of at least 10,000 (but less than 50,000) population. Each metro or micro area consists of one or more counties and includes the counties containing the core urban area, as well as any adjacent counties that have a high degree of social and economic integration (as measured by commuting to work) with the urban core.

**Coverage.** A measure of how well a frame and sample design includes the universe of interest. Coverage is typically expressed as a proportion. For example, if a study has 75% coverage, then the frame and the sample design include 75% of the universe of interest.

**Crime.** Victimizations and incidents are classified into crime categories based on detailed characteristics of the event provided by the respondent. The classifications include personal crimes, violent crimes, and property crimes. See “Incident” and “Victimization.”

**Personal crime.** Rape, sexual assault, personal robbery, aggravated and simple assault, purse snatching, and pocket picking. This category includes both attempted and completed crimes.

**Violent crime.** Rape, sexual assault, robbery, aggravated assault, and simple assault. This category includes both attempted and completed crimes but excludes purse snatching and pocket picking. Murder is not measured by the NCVS because of the inability to question the victim. Completed violence refers to the sum of all completed rapes, sexual assaults, robberies, aggravated assaults, and simple assaults. Attempted violence refers to the unsuccessful attempt of rape, sexual assault, personal robbery, and assault. Threats of harm include attempted attacks and attempted sexual assaults by means of verbal threats.

**Property crime.** Burglary, motor vehicle theft, and theft. This category includes both attempted and completed crimes.

For more information on specific crimes, see the terms and definitions on the BJS website <<http://www.bjs.gov>>.

**Data year.** The set of victimizations reported to NCVS that occurred within the same calendar year. For all of the years prior to 1996, tables on criminal victimization in the United States are based on data year. Since 1996, tables have been based on collection year. See "Collection Year."

**Dependent listing.** A listing that adds, subtracts, and revises the information on a prior list. See “Independent Listing.”

**Domain of interest, or domain.** A specific subset of the universe.

**Eligible/Ineligible** – Whether a unit of interest is in the universe of interest or not in the universe of interest. See also AAPOR (2011).

**Frame.** The list of units in the universe of interest.

**Field representative (FR)** – The Census Bureau’s term for an interviewer.

**Generalized Variance Function (GVF).** A simple model that expresses the variance as a function of the expected value of the survey estimate (Wolter, 1984).

**Group Quarters (GQ).** According to the 2010 Census—

*"A GQ is a place where people live or stay that is normally owned or managed by an entity or organization providing housing and/or services for the residents. These services may include custodial or medical care as well as other types of assistance, and residency is commonly restricted to those receiving these services. People living in GQs are usually not related to each other. GQs 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."*

**Half sample.** This is a code that identifies one of the two PSUs within a pseudo stratum.

**Household.** Address, House or GQ, and housing unit Equivalent. See “Unit.”

**Housing Unit (HU)** –A house, apartment, mobile home or trailer, a group of rooms, or a single room occupied as separate living quarters, or if vacant, intended for occupancy as separate living quarters. Separate living quarters are those in which the occupants live separately from any other individuals in the building and have direct access from outside the building or through a common hall. For vacant units, the criteria of separateness and direct access are applied to the intended occupants whenever possible.

**Incident.** A specific criminal act involving one or more victims and offenders. For example, if two persons are robbed at the same time and place, this crime is classified as two robbery victimizations and one robbery incident.

**Independent listing.** A listing that starts with no prior list. See “Dependent listing.”

**Listing.** The general term for the identification of units. Listing can occur in the geographic area of interest, a building permit office, or within a specific GQ. When a unit is identified, the unit is put on a list that can be used as the sampling frame for the survey.

**Measure of size (MOS).** This is a quantity used in unequal sample selection methods to define the probabilities of selection. The MOS is important because sample designs with probabilities of selection that are proportional to the variable of interest can have small variances. If the MOS is exactly proportional to the variable of interest, then the sampling variance will be zero.

**Noninterview.** Eligible units are classified as either completed interviews or noninterviews. Because the unit of interest for NCVS is both households and persons, the NCVS has both household and person noninterviews.

**Nonresponse.** The two basic types of nonresponse are unit nonresponse and item nonresponse.

**Permit frame.** A list of addresses for all of the units built since the most recent decennial census. Initially, the permit frame is an empty list of placeholders that are used to select the sample. During the decade, permit counts from specific Building Permit Offices (BPO) and dates are matched to the permit frame, which identifies BPOs that include selected sample units. Field staff list the identified BPOs, and the final sample is identified from the listing.

**Primary sample unit (PSU).** The first-stage unit of a multi-stage sample design. PSU is a large metropolitan area, county, or group of bordering counties.

**Pseudo strata.** Sometimes called variance strata because they are used in variance estimation or standard error computational unit (SECU) codes. Because NCVS selects one PSU per first-stage stratum, the direct estimator of the variance is unbiased. To estimate the variance, the first-stage strata are paired, with two PSUs contained within each strata.

**Reference period.** The period for which the respondent is asked to report characteristics of interest. For NCVS, the reference period is the 6 months prior to the day of the interview.

**Reference person.** For NCVS, the reference person is a responsible adult household member who is not likely to permanently leave the household. To meet these criteria, the reference person must—

- qualify as a household member at the sample address
- be one of the owners or renters at the sample address
- be at least age 18 (in most cases).

Because an owner or renter of the sample housing unit is normally the most responsible and knowledgeable household member, this person is generally designated as the reference person and household respondent. However, a household respondent does not have to be one of the household members who owns or rents the unit.

**Relative variance, relvariance, or relvar.** This is a measure of the relative dispersion of a probability distribution and is defined as the variance divided by the square of the estimate. It is also equal to the square of the coefficient of variation. That is,  $\sqrt{v(\hat{\theta})}/\hat{\theta}$ .

**Sample design.** This includes everything about the selection of units in the sample that determines the probability of selection for each unit. Estimation is considered separate from sample design, in that some estimation procedures are more appropriate than others for a given sample design, but any estimator could be used with the sample derived from a given sample design.

**Sampling fraction.** The fraction of the universe that is in the sample. With an equal probability sample design, the sampling fraction is the ratio of the sample size to the size of the universe, often represented as  $f = n/N$ .

**Sampling interval.** The inverse of the sampling fraction. It is sometimes referred to as the “take-every” because every  $f^{-1}$  units of the universe is selected in the sample.

**Self-representing/Non self-representing (SR/NSR).** A unit is self-representing (SR) if its probability of selection is 1.0, and a unit is non self-representing (NSR) if its probability of selection is other than 1.0. A unit that is SR represents itself and no other PSUs because it is the only PSU in its stratum. A unit that is NSR represents itself and the other units of the same stratum. The terms certainty and non-certainty are used the same way as SR and NSR, respectively.

**Self-weighting.** A type of sample design in which units have equal probabilities of selection. NCVS has obtained equal overall probabilities of selection by compensating probabilities at different stages. Many household surveys are self-weighting because not much is known about

specific households prior to interviewing. These household are considered equally important as contributing to the estimate. See also Kish (1965, p. 221).

**Special place.** A special place is an administrative entity containing one or more group quarters where persons live or stay, such as a college or university, nursing home, hospital, correctional facility, or military installation or ship. A special place can include one or more GQs. HUs can also be a special place (e.g., a college president's home on a college campus).

**srswor.** Simple random sampling without replacement. In sampling without replacement, each sample unit of the population has only one chance to be selected in the sample. A population of  $N$  units can be thought of drawing  $n$  independent samples of size 1. One unit is randomly selected from the population to be the first sampled unit, with probability  $1/N$ . A second unit is randomly selected with probability  $1/(N-1)$  since the first unit is not replaced back into the population. This procedure is repeated until the sample has  $n$  units, which, by design, cannot include duplicates from the population.

**srswr.** Simple random sampling with replacement. In sampling with replacement, a population of  $N$  units can be thought of drawing  $n$  independent samples of size 1. One unit is randomly selected from the population to be the first sampled unit, with probability  $1/N$ . Then the sampled unit is replaced in the population, and a second unit is randomly selected with probability  $1/N$ . This procedure is repeated until the sample has  $n$  units, which may include duplicates from the population.

**Stratified sampling.** A sample design that partitions the universe of interest into strata and selects an independent sample from each stratum. "If intelligently used, stratification nearly always results in smaller variance for the estimated mean or total than is given by a comparable simple random sample" (Cochran, 1977, p. 99).

**sys.** Systematic random sampling from an ordered list.

**Successive difference replication (SDR).** A replication variance estimation method that mimics the successive difference variance estimator and can be used to estimate the variance from a *sys* sample design. The main ideas of replication are outline by Fay and Train (1995).

**Time-in-sample (TIS).** The interview number of a given unit. For example, a unit in its third interview is also referred to as TIS 3, or the unit's third time-in-sample.

**Unit.** According to Hájek (1981, p. 4)—

*The units making up the population  $S$  may be any elements worth studying—persons, families, farms, account items, temperature readings, and so on—and their nature will be irrelevant for theoretical considerations. The units are assumed to be identifiable by certain labels (tags, names, addresses) and that are available in the frame (list, map) showing how to reach any unit given its label.*

For NCVS, unit will make no distinctions between household, address, house, housing unit, housing unit equivalent, or group quarters. Important operational distinctions between all of these terms exist. However, for purpose of describing the methodology of NCVS, the general term "unit" is examined for ease of presentation, unless the distinction is necessary.

**Unit frame.** A list of addresses representing all of the housing units that exist prior to census day (April 1) for the prior decennial census. For example, the unit frame for the Census 2000 design includes all of the addresses built before April 1, 2000 compiled for the Census 2000.

**Universe of interest.** In finite population sampling, the universe of interest, or simply the universe, is the well-defined set of units for which an estimate will be generated.

**Urbanicity/location of residence.** The Census Bureau's urban-rural classification is fundamentally a delineation of geographical areas, identifying both individual urban areas and the rural areas of the nation. The Census Bureau's urban areas represent densely developed territory, and encompass residential, commercial, and other non-residential urban land uses. The Census Bureau delineates urban areas after each decennial census by applying specified criteria to decennial census and other data. The Census Bureau identifies two types of urban areas: Urbanized Areas (UAs) of 50,000 or more people; and Urban Clusters (UCs) of at least 2,500 and less than 50,000 people. "Rural" encompasses all population, housing, and territory not included within an urban area.

**Variance or sample variance.** A measure of the variability of an estimate. With finite population sampling, variance refers to the measure of how the estimate may differ if other samples were selected. Formally, the variance of an estimator  $\hat{\theta}$  is the expected value of the squared difference between the estimator  $\hat{\theta}$  and its expected value. That is,  $v(\hat{\theta}) = E(\hat{\theta} - E(\hat{\theta}))^2$ .

**Victimization.** A crime as it affects one individual person or household. For personal crimes, the number of victimizations is equal to the number of victims involved. The number of victimizations may be greater than the number of incidents because more than one person may be victimized during an incident. Each crime against a household is assumed to involve a single victim—the affected household.

## Appendix B: 2013 monthly coverage ratios

**Appendix table B1. Coverage ratios in January 2013**

Demographic characteristic	Coverage ratio
Sex	
Male	0.82
Female	0.84
Race/Hispanic origin	
White	0.82
Black/African American	0.86
Hispanic/Latino	0.78
American Indian/Alaska Native	0.53
Asian/Native Hawaiian/Other Pacific Islander	0.73
Two or more races	0.61
Age	
12–17	0.79
18–24	0.66
25–34	0.80
35–49	0.84
50–64	0.86
65 or older	0.95

**Appendix table B2. Coverage ratios in February 2013**

Demographic characteristic	Coverage ratio
<b>Sex</b>	
Male	0.81
Female	0.83
<b>Race/Hispanic origin</b>	
White	0.76
Black/African American	0.85
Hispanic/Latino	0.76
American Indian/Alaska Native	0.52
Asian/Native Hawaiian/Other Pacific Islander	0.91
Two or more races	0.66
<b>Age</b>	
12–17	0.76
18–24	0.67
25–34	0.77
35–49	0.84
50–64	0.87
65 or older	0.90

**Appendix table B3. Coverage ratios in March 2013**

Demographic characteristic	Coverage ratio
Sex	
Male	0.84
Female	0.84
Race/Hispanic origin	
White	0.81
Black/African American	0.85
Hispanic/Latino	0.83
American Indian/Alaska Native	0.60
Asian/Native Hawaiian/Other Pacific Islander	0.91
Two or more races	0.69
Age	
12–17	0.76
18–24	0.73
25–34	0.84
35–49	0.85
50–64	0.88
65 or older	0.91

**Appendix table B4. Coverage ratios in April 2013**

Demographic characteristic	Coverage ratio
Sex	
Male	0.84
Female	0.88
Race/Hispanic origin	
White	0.84
Black/African American	0.90
Hispanic/Latino	0.74
American Indian/Alaska Native	0.51
Asian/Native Hawaiian/Other Pacific Islander	0.82
Two or more races	0.69
Age	
12–17	0.87
18–24	0.69
25–34	0.80
35–49	0.87
50–64	0.94
65 or older	0.91

**Appendix table B5. Coverage ratios in May 2013**

Demographic characteristic	Coverage ratio
Sex	
Male	0.84
Female	0.85
Race/Hispanic origin	
White	0.83
Black/African American	0.87
Hispanic/Latino	0.76
American Indian/Alaska Native	0.85
Asian/Native Hawaiian/Other Pacific Islander	0.79
Two or more races	0.59
Age	
12–17	0.76
18–24	0.71
25–34	0.85
35–49	0.86
50–64	0.87
65 or older	0.91

**Appendix table B6. Coverage ratios in June 2013**

Demographic characteristic	Coverage ratio
Sex	
Male	0.82
Female	0.85
Race/Hispanic origin	
White	0.86
Black/African American	0.86
Hispanic/Latino	0.80
American Indian/Alaska Native	0.77
Asian/Native Hawaiian/Other Pacific Islander	0.66
Two or more races	0.69
Age	
12–17	0.80
18–24	0.68
25–34	0.77
35–49	0.84
50–64	0.89
65 or older	0.94

**Appendix table B7. Coverage ratios in July 2013**

Demographic characteristic	Coverage ratio
Sex	
Male	0.81
Female	0.84
Race/Hispanic origin	
White	0.82
Black/African American	0.85
Hispanic/Latino	0.72
American Indian/Alaska Native	0.51
Asian/Native Hawaiian/Other Pacific Islander	0.81
Two or more races	0.66
Age	
12–17	0.77
18–24	0.63
25–34	0.80
35–49	0.83
50–64	0.85
65 or older	0.97

**Appendix table B8. Coverage ratios in August 2013**

Demographic characteristic	Coverage ratio
Sex	
Male	0.82
Female	0.84
Race/Hispanic origin	
White	0.77
Black/African American	0.85
Hispanic/Latino	0.79
American Indian/Alaska Native	0.43
Asian/Native Hawaiian/Other Pacific Islander	0.93
Two or more races	0.60
Age	
12–17	0.79
18–24	0.68
25–34	0.77
35–49	0.85
50–64	0.87
65 or older	0.92

**Appendix table B9. Coverage ratios in September 2013**

Demographic characteristic	Coverage ratio
Sex	
Male	0.84
Female	0.85
Race/Hispanic origin	
White	0.76
Black/African American	0.87
Hispanic/Latino	0.79
American Indian/Alaska Native	0.66
Asian/Native Hawaiian/Other Pacific Islander	0.87
Two or more races	0.72
Age	
12–17	0.76
18–24	0.68
25–34	0.84
35–49	0.86
50–64	0.88
65 or older	0.93

**Appendix table B10. Coverage ratios in October 2013**

Demographic characteristic	Coverage ratio
Sex	
Male	0.77
Female	0.81
Race/Hispanic origin	
White	0.77
Black/African American	0.83
Hispanic/Latino	0.64
American Indian/Alaska Native	0.53
Asian/Native Hawaiian/Other Pacific Islander	0.76
Two or more races	0.62
Age	
12–17	0.75
18–24	0.61
25–34	0.75
35–49	0.79
50–64	0.87
65 or older	0.87

**Appendix table B11. Coverage ratios in November 2013**

Demographic characteristic	Coverage ratio
Sex	
Male	0.86
Female	0.87
Race/Hispanic origin	
White	0.80
Black/African American	0.90
Hispanic/Latino	0.78
American Indian/Alaska Native	0.84
Asian/Native Hawaiian/Other Pacific Islander	0.86
Two or more races	0.62
Age	
12–17	0.79
18–24	0.70
25–34	0.84
35–49	0.87
50–64	0.92
65 or older	0.94

**Appendix table B12. Coverage ratios in December 2013**

Demographic characteristic	Coverage ratio
Sex	
Male	0.88
Female	0.89
Race/Hispanic origin	
White	0.95
Black/African American	0.91
Hispanic/Latino	0.81
American Indian/Alaska Native	0.76
Asian/Native Hawaiian/Other Pacific Islander	0.75
Two or more races	0.67
Age	
12–17	0.89
18–24	0.77
25–34	0.82
35–49	0.88
50–64	0.94
65 or older	0.98

## Appendix C: GVF parameters for NCVS

**Appendix table C1. NCVS GVF parameter set, 2013**

<b>GVF</b>	<b><i>a</i></b>	<b><i>b</i></b>	<b><i>c</i></b>
Overall person crime estimates	-0.00082347	2059	13.264
Person crime domain estimates	-0.00111800	1958	18.059
Overall property crime estimates	-0.00023036	2084	2.418
Property crime domain estimates	-0.00021207	2786	2.149

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2013.

**Appendix table C2. NCVS GVF parameter set, 2012**

<b>GVF</b>	<b><i>a</i></b>	<b><i>b</i></b>	<b><i>c</i></b>
Overall person crime estimates	-0.00025343	6439	3.704
Person crime domain estimates	-0.00054187	3893	8.530
Overall property crime estimates	-0.00020722	2345	2.116
Property crime domain estimates	-0.00020565	2810	2.057

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2012.

**Appendix table C3. NCVS GVF parameter set, 2011**

<b>GVF</b>	<b><i>a</i></b>	<b><i>b</i></b>	<b><i>c</i></b>
Overall person crime estimates	-0.00060211	2439	9.511
Person crime domain estimates	-0.00081383	2309	12.916
Overall property crime estimates	-0.00028723	4182	2.809
Property crime domain estimates	-0.00037148	2981	3.852

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2011.

**Appendix table C4. NCVS GVF parameter set, 2010**

<b>GVF</b>	<b><i>a</i></b>	<b><i>b</i></b>	<b><i>c</i></b>
Overall person crime estimates	-0.00086145	1384	13.696
Person crime domain estimates	-0.00108760	1960	17.278
Overall property crime estimates	-0.00020353	2450	2.035
Property crime domain estimates	-0.00017899	3297	1.687

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2010.

**Appendix table C5. NCVS GVF parameter set, 2009**

<b>GVF</b>	<b><i>a</i></b>	<b><i>b</i></b>	<b><i>c</i></b>
Overall person crime estimates	-0.00054111	2410	8.475
Person crime domain estimates	-0.00106740	3529	16.794
Overall property crime estimates	-0.00030998	2160	3.233
Property crime domain estimates	-0.00026607	3098	2.663

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2009.

**Appendix table C6. NCVS GVF parameter set, 2008**

<b>GVF</b>	<b><i>a</i></b>	<b><i>b</i></b>	<b><i>c</i></b>
Overall person crime estimates	-0.00063195	5004	9.722
Person crime domain estimates	-0.00076651	5025	11.857
Overall property crime estimates	-0.00035885	1887	3.778
Property crime domain estimates	-0.00032519	3348	3.275

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2008.

## Appendix D: GVF parameters for NCVS supplements

**Appendix table D1. GVF parameter set for NCVS supplements, 1999—2013**

<b>GVF</b>	<b><i>a</i></b>	<b><i>b</i></b>	<b><i>c</i></b>
2013 SCS	-0.00109539	4,721	5.066
2012 ITS	-0.00064099	4,852	9.755
2011 PPCS	-0.00051702	3,115	7.845
2011 SCS	-0.00020076	3,060	3.028
2009 SCS	-0.00043149	3,465	4.490
2008 PPCS	-0.00032181	4,252	3.058
2008 ITS	-0.00040650	5,441	3.843
2006 SVS	-0.00034318	3,594	3.434
2005 PPCS	-0.00062481	3,789	6.215
2003 SCS	-0.00029301	3,059	2.872
2002 PPCS	-0.00028715	2,589	2.677
2001 SCS	0.00011330	2,803	2.905
1999 SCS	-0.00026646	2,579	2.826
1999 PPCS	-0.00027776	2,568	2.537

SCS – School Crime Supplement.

PPCS – Police Public Contact Survey.

ITS – Identity Theft Supplement.

SVS – Supplemental Victimization Survey.

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 1999—2013.

## Appendix E: Calculating replicate variances with SAS

### NCVS files needed

The public use and replicate weight files are located on the National Archive of Criminal Justice Data (NACJD) website <<http://www.icpsr.umich.edu/icpsrweb/NACJD/NCVS/>>.

The variable names on the public use files will differ from what the Census Bureau uses, but the layout of the person-level replicate weight file may be presented as follows:

**Appendix table E1. Person-level replicate weight file, by variable and description**

Variable	Description
PERWGT	Full sample weight
PERWGT1	Replicate weight 1
PERWGT2	Replicate weight 2
PERWGT3	Replicate weight 3
⋮	⋮
PERWGT(n)	Replicate weight n
⋮	⋮
PERWGT160	Replicate weight 160
IDPER	Person identification number
YEARQ	Year and quarter

The variables IDPER and YEARQ will be used to match to the NCVS data.

### Merging files

To merge with the NCVS person or household file, sort the file of interest and the corresponding NCVS replicate weight file by IDPER (or IDHH for household-level data) and YEARQ, and then merge the files.

The following SAS code includes an example on how to successfully merge NCVS files to the NCVS replicate weights file.

### SAS code to perform calculations

The following is an example on how to calculate variances using the NCVS person replicate weight file. Again, the variable and file names may not match what is listed on the public use files.

```
*****;
* The FIRST STEP is to sort the data files by  *;
* IDPER YEARQ.                               *;
*****;
proc sort data=NCVSPersfile; by IDPER YEARQ; run;
proc sort data=NCVSRWfile; by IDPER YEARQ; run;
```

```

*****;
* Next, merge the two files above.    *;
* This example flags persons who are black/African American females  *;
*****;
data user.data1;
merge NCVSPersfile NCVSRWfile;
by IDPER YEARQ;
run;
*****;
* For our example, we want to estimate the variance of  *;
* black/African American females, so we flag persons who are black/African American
females.*;
*****;
data user.data1;
set user.data1;
if racep=2 and sexp = 2 then bfem=1;
else bfem = 0;
run;
*****;
* Next we want to sum the full sample and the  *;
* 160 replicate weights and write them out to a file. *;
*****;
proc means data=user.data1 sum noprint;
where bfem=1;
var perwgt perwgt1-perwgt160;
output out=user.data2 sum=est rw1-rw160;
run;
proc print data = user.data2;
var est rw1-rw160;
run;
*****;
* The THIRD STEP of code uses the estimates of the full *;
* sample and the 160 replicates to compute the estimated *;

```

```

* replicate variance(s) using the formula(s) for 160 *;
* replicates.      *;
*****;

data user.data3 (keep=char est var se cv);
set user.data2 end=eof;
if _n_=1 then sdiffsq = 0;
array repwts{ 161 } est rw1-rw160;
do I = 2 to 161;
sdiffsq = sdiffsq + (repwts{I} - repwts{1})**2;
end;
if eof then do;
var = (4/160) * sdiffsq;
length char $25;
char = 'Variance of Black Females';
output;
end;
run;
proc print data = user.data3;
var char est var;
run;

```

## Appendix F: 1-year correlations

**Appendix table F1. NCVS 1-year correlations for personal crime, 1999—2013**

	2012–2013	2011–2012	2010– 2011	2009– 2010	1999– 2009/a
Total personal crime	0.32	0.10	0.20	0.25	0.30
Violent crime	0.33	0.11	0.20	0.24	0.31
Completed	0.13	0.08	0.03	0.02	0.19
Attempted	0.30	0.06	0.14	0.30	0.29
Rape/sexual assault	0.06	0.05	-0.007	-0.04	0.04
Rape	0.06	0.03	-0.03	0.02	0.03
Completed	-0.04	0.03	-0.17	0.007	0.06
Attempted	0.08	-0.07	-0.03	0.02	-0.01
Sexual assault	-0.001	-0.05	0.01	-0.04	0.09
Robbery	0.21	0.11	0.09	0.14	0.04
Completed	0.22	0.13	0.07	0.27	0.00
With injury	-0.01	0.10	0.17	0.07	-0.03
Without injury	0.03	0.14	0.007	0.15	0.02
Attempted	-0.01	-0.11	-0.10	0.12	0.01
With injury	-0.03	0.02	0.05	-0.18	0.10
Without injury	-0.005	-0.15	-0.16	0.18	0.01
Assault	0.28	0.14	0.14	0.15	0.30
Aggravated	0.003	0.10	0.14	0.05	0.09
Completed with injury	0.02	0.12	0.06	0.07	0.07
Attempted/threatened with weapon	-0.02	0.04	0.13	0.10	0.07
Simple	0.28	0.05	0.12	0.12	0.26
Completed with injury	-0.06	-0.02	0.002	0.04	0.15
Attempted/threatened without weapon	0.33	-0.003	0.15	0.14	0.22
Purse snatching/pocket picking/b	-0.05	-0.01	-0.06	-0.05	0.03
Completed	-0.03	0.002	-0.02	-0.18	0.04
Attempted	0.02	-0.01	0.00	0.00	0.00

a/From 1999 to 2009, the same correlations were used every year. Beginning in 2010, the correlation formula was revised to calculate new correlations annually.

b/Users who wish to make estimates for either purse snatching or pocket picking are encouraged to use the correlation provided under purse snatching/pocket picking.

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 1999–2013.

**Appendix table F2. NCVS 1-year correlations for property crime, 1999–2013**

	<b>2012– 2013</b>	<b>2011– 2012</b>	<b>2010– 2011</b>	<b>2009– 2010</b>	<b>1999– 2009*</b>
Total property crime	0.40	0.47	0.35	0.42	0.38
Burglary	0.26	0.14	0.11	-0.02	0.21
Completed	0.23	0.13	0.14	-0.02	0.17
Forcible entry	0.12	0.15	-0.07	-0.03	0.15
Unlawful entry without force	0.13	0.06	0.15	0.12	0.12
Attempted forcible entry	0.01	0.02	-0.10	-0.20	0.09
Motor vehicle theft	0.09	0.21	-0.06	-0.08	0.08
Completed	0.04	0.07	-0.008	0.006	0.05
Attempted	0.14	0.20	-0.07	0.06	0.04
Theft	0.28	0.47	0.38	0.36	0.34
Completed	0.27	0.44	0.35	0.33	0.34
Less than \$50	0.10	0.06	0.11	0.05	0.22
\$50–\$249	0.15	0.20	0.15	0.22	0.20
\$250 or more	0.21	0.24	0.05	0.09	0.19
Amount not available	0.07	0.22	0.18	0.11	0.09
Attempted	0.01	-0.07	0.10	-0.02	0.08

\*From 1999 to 2009, the same correlations were used every year. Beginning in 2010, the correlation formula was revised to calculate new correlations annually.

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 1999–2013.

## Appendix G: 2-year correlations

**Appendix table G1. NCVS 2-year correlations for personal crime, 1999–2013**

	2011–2013	2010–2012	2009–2011	2008– 2010	1999– 2009/a
Total personal crime	0.05	-0.10	0.14	0.32	0.15
Violent crime	0.07	-0.08	0.14	0.31	0.15
Completed	0.09	-0.03	0.17	0.22	0.09
Attempted	0.02	-0.17	0.11	0.26	0.14
Rape/sexual assault	0.09	0.06	0.66	0.01	0.02
Rape	0.01	0.05	0.30	0.04	0.01
Completed	-0.04	0.02	0.10	0.11	0.03
Attempted	0.05	0.04	0.25	0.03	0.00
Sexual assault	0.09	0.19	0.01	0.02	0.04
Robbery	0.08	0.17	-0.04	0.31	0.02
Completed	0.05	0.16	0.03	0.39	0.00
With injury	0.03	0.002	-0.10	0.10	-0.01
Without injury	0.12	0.18	0.15	0.40	0.01
Attempted	-0.06	0.005	0.006	0.24	0.00
With injury	0.04	-0.02	0.01	-0.08	0.05
Without injury	-0.08	0.02	-0.02	0.22	0.00
Assault	-0.002	-0.15	0.12	0.30	0.15
Aggravated	0.01	0.03	0.21	0.03	0.04
Completed with injury	0.09	0.03	0.14	-0.22	0.03
Attempted/threatened with weapon	0.17	-0.003	0.08	0.17	0.03
Simple	-0.08	-0.16	0.14	0.22	0.13
Completed with injury	-0.02	-0.12	0.11	0.06	0.07
Attempted/threatened without weapon	-0.08	-0.19	0.13	0.19	0.11
Purse snatching/pocket picking/b	-0.19	-0.07	0.05	-0.01	0.01
Completed	-0.01	0.02	0.07	0.04	0.02
Attempted	0.02	0.00	-0.04	0.00	0.00

a/From 1999 to 2009, the same correlations were used every year. Beginning in 2010, the correlation formula was revised to calculate new correlations annually.

b/Users who wish to make estimates for either purse snatching or pocket picking are encouraged to use the correlation provided under purse snatching/pocket picking.

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 1999–2013.

**Appendix table G2. NCVS 2-year correlations for property crime, 1999–2013**

	<b>2011– 2013</b>	<b>2010– 2012</b>	<b>2009– 2011</b>	<b>2008– 2010</b>	<b>1999–2009</b>
Total property crime	0.42	0.50	0.19	0.33	0.19
Burglary	0.26	0.15	0.02	0.32	0.10
Completed	0.27	0.23	0.04	0.33	0.08
Forcible entry	0.12	0.10	-0.08	0.10	0.07
Unlawful entry without force	0.14	0.17	0.01	0.33	0.06
Attempted forcible entry	-0.19	-0.08	0.01	0.07	0.04
Motor vehicle theft	0.19	0.09	0.09	0.07	0.04
Completed	0.23	0.03	0.07	0.07	0.02
Attempted	0.16	0.05	0.02	0.16	0.02
Theft	0.32	0.53	0.29	0.31	0.17
Completed	0.34	0.53	0.32	0.28	0.17
Less than \$50	0.20	0.11	-0.10	0.06	0.11
\$50–\$249	0.20	0.36	0.14	0.03	0.10
\$250 or more	0.12	0.19	0.19	0.10	0.09
Amount not available	0.01	0.19	0.19	0.15	0.04
Attempted	-0.17	0.15	-0.11	-0.04	0.04

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 1999–2013.

## Appendix H: Point and variance estimation procedures

Several different weights are produced for NCVS that are specialized for estimating specific types of totals. This section provides a brief review of each of the weights, including the purpose of each and how each weight differs from other weights. Note that subsequent sections provide more information about the composition of weights.

The four data files for NCVS are provided at four different levels, including—

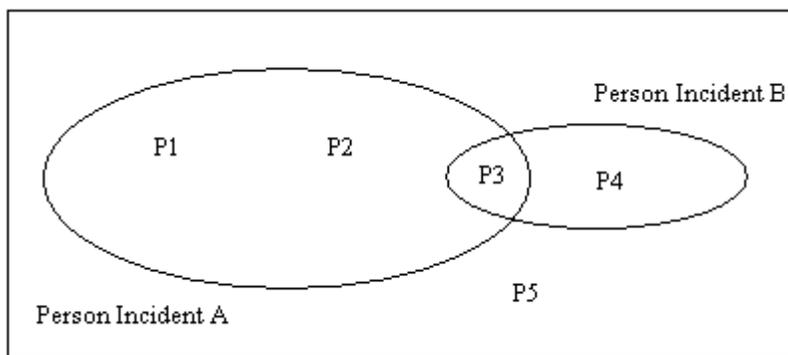
- address file
- household file
- person file
- incident (or victimization) file.

These files are used to generate estimates of the total household and person populations for a given year and the number of criminal incidents and victimizations over the same period. Household weights are on the household file, and person weights are on the person file. Note that every household interviewed or not interviewed is on the household file.

The incident file will contain an observation for each victimization (i.e., person/incident combination). A person may show up multiple times on the incident file if he or she experienced multiple victimizations. The file has one observation for each person who experienced an incident. The incident file has both a victimization weight and an incident weight. The victimization weight is used to count the number of unique victimizations. To prevent double counting of incidents, the incident weight is equal to the victimization weight divided by the number of victims in the incident. One incident may have multiple victimizations (i.e., more than one victim), and a person may suffer multiple incidents and victimizations over time.

To understand how the incident and victimization weights relate, consider the following simple example of two incidents involving a total of five different persons.

### Appendix figure H.1. Representation of NCVS person victimizations and incidents



Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2013.

In appendix figure H.1, the five persons in our example are identified as P1, P2, P3, P4, and P5. The two ovals represent two separate incidents. In incident A, three persons P1, P2, and P3 were involved in a robbery that happened at the same time. In incident B, persons P3 and P4 were involved in a single incident that differs from incident A. Person P5 was in the survey but was

not victimized. To summarize, four unique persons were victimized in two different incidents, with a total of five victimizations.

The person file for this example would have an observation for each of the five persons in the survey, P1 through P5.

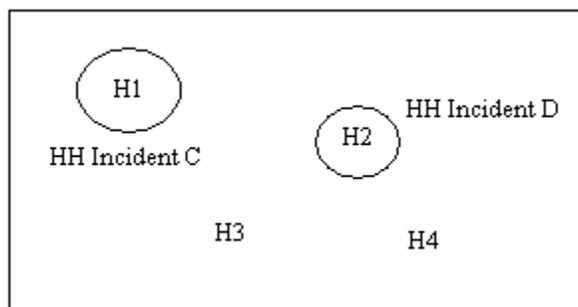
**Appendix table H.1. Representation of the NCVS person file**

<b>Person</b>	<b>Person weight</b>
P1	<i>PERSONWGT</i> <sub>P1</sub>
P2	<i>PERSONWGT</i> <sub>P2</sub>
P3	<i>PERSONWGT</i> <sub>P3</sub>
P4	<i>PERSONWGT</i> <sub>P4</sub>
P5	<i>PERSONWGT</i> <sub>P5</sub>

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2013.

Appendix figure H.2 shows a representation of households and household incidents. In general, household incidents are much simpler to calculate than person incidents because only one household can be in a given household incident. As presented in appendix figure H.2, household H1 was involved in incident C, and household H2 was involved in incident D. Households H3 and H4 were not involved in any incidents.

**Appendix figure H.2. Representation of NCVS household victimizations and incidents**



Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2013.

Appendix table H.2 summarizes the households and household weights.

**Table H.2. Representation of the NCVS household file**

<b>Household</b>	<b>Household weight</b>
H1	<i>HHWGT</i> <sub>H1</sub>
H2	<i>HHWGT</i> <sub>H2</sub>
H3	<i>HHWGT</i> <sub>H3</sub>
H4	<i>HHWGT</i> <sub>H4</sub>

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2013.

Appendix table H.3 shows that one record is created for each reported person and household victimization in the incident file.

**Appendix table H.3: Representation of the NCVS incident file**

<b>Incident</b>	<b>Household or person victimization</b>	<b>Victimization weight</b>	<b>Incident weight</b>
A	P1	<i>VWGT</i> <sub>P1,A</sub>	<i>INCIDENTWGT</i> <sub>P1,A</sub>
A	P2	<i>VWGT</i> <sub>P2,A</sub>	<i>INCIDENTWGT</i> <sub>P2,A</sub>
A	P3	<i>VWGT</i> <sub>P3,A</sub>	<i>INCIDENTWGT</i> <sub>P3,A</sub>
B	P3	<i>VWGT</i> <sub>P3,B</sub>	<i>INCIDENTWGT</i> <sub>P3,B</sub>
B	P4	<i>VWGT</i> <sub>P4,B</sub>	<i>INCIDENTWGT</i> <sub>P4,B</sub>
C	H1	<i>VWGT</i> <sub>H1,B</sub>	<i>INCIDENTWGT</i> <sub>H1,B</sub>
D	H2	<i>VWGT</i> <sub>H2,B</sub>	<i>INCIDENTWGT</i> <sub>H2,B</sub>

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2013.

The file has one observation for each person that was part of incident A and one observation for each person that was part of incident B. The file has two observations for person P3 because he or she was victimized twice. The incident weight is equal to the victimization weight divided by the number of victims in the incident. If the only victims of incident A were persons P1, P2, and P3, their incident weight would be a third of their victimization weight. Because the incident weight is divided by three, when the three victimizations of incident A are summed, a total of one incident is generated.

Note that person P5 is not in the incident file. Although he or she was interviewed, person P5 is excluded from the file because he or she was not victimized.

Household victimizations are simple to calculate because each victimization only includes one household. Only households victimized are included in the incident file.

The organization of the incident file allows for the greatest amount of flexibility in producing estimates. Estimates can be made of the number of victimizations or incidents in different domains in which domains are defined in terms of the characteristics of the incident, characteristics of the person or household victimized, or characteristics of both the incident and the victimized. For example, the domain can be defined as those incidents that occurred in the evening and involved persons with at least a college degree. The evening is a characteristic of the incident, and education is a characteristic of the persons in the incident.

## How to estimate statistics with the survey weights

This section discusses the following topics related to weighting and estimation:

- A review of the sample design weights
- How to use the weights to estimate different types of collection-year statistics.

**Household weights (*HHWEIGHT*).** The household weight can be used to estimate the number of households with a characteristic ( $N_{HHs}$ ) or the total of some household-level variable of interest ( $Y_{HHs}$ ).

**Person weights (*PERSONWEIGHT*).** The person weight can be used to estimate the number of persons with a characteristic ( $N_{persons}$ ) or the total of some person-level variable of interest ( $Y_{persons}$ ).

If everyone in the household responded to the NCVS, the person weight would be the same as the household weight. Because persons within households do not always respond to survey requests, noninterviews are accounted for with a person-level nonresponse factor.

**Victimization weights (*VWGT*).** The victimization weight can be used to estimate the number of household-level victimizations  $\left(N_{HH\text{ victimizations}}\right)$  and the number of person-level victimizations

$\left(N_{person\text{ victimizations}}\right)$ . The victimization weight incorporates the bounding adjustment.

In 2007, NCVS started using the first interview of the seven NCVS interviews in estimates. Prior to 2007, the first interview was a bounding interview and not used in the estimates. Because NCVS has a known time-in-sample effect (i.e., respondents in earlier interviews report more incidents than the same respondents in later interviews), adding the first interview without an adjustment would have increased victimization rates, compared to prior survey years.

To make the data from the seven interviews comparable to data from six interviews, a bounding adjustment is applied to the weights of units in the first interview. Because the first interview has the most reported incidents, the adjustment reduces the reporting of the first interview. The adjustment is calculated and applied separately for household incidents and person incidents. (See Chapter 5 for more information on bounding adjustment.) Prior to 2007, the household weight or person weight was applied to incident records to estimate weighted incidents.

Incidents are associated with both households and persons. For example, burglary is associated with households, and assault is associated with a person. Victimization weights are put directly on the victimizations, either household or person, listed in the incident file.

**Incident weights (*INCIDENTWEIGHT*).** The incident weight can be used to estimate the number of household incidents  $\left(N_{HH\text{ incidents}}\right)$  and the number of person incidents  $\left(N_{person\text{ incidents}}\right)$ .

Incidents are associated with both households and persons in the same way that victimizations are associated with both households and persons.

If every incident only had one victim, the incident weight would be the same as the victimization weight. Because incidents can have multiple victims, each incident weight is adjusted specific to a given incident to account for multiple victims of a given incident. (See the section on the multiple victim factor for more information on this adjustment for persons.) For households, the victimization weight is the same as the incident weight.

**Estimators of household-level totals.** The estimator for  $N_{HH}$ , the total number of households in a domain of interest, is the sum of the household weights for all sample households in the domain of interest within a 6-month period. That is,

$$\hat{N}_{HH} = \sum_{\substack{\text{sample} \\ \text{HHs} \\ \text{in 6 months}}} HHWEIGHT_i$$

where  $i$  is the index over the sum.

The sum of the *HHWEIGHT* for all of the completed household interviews over 6 consecutive months is an estimate of the number of households in the United States. If you sum a year of completed interviews, the total should be adjusted by an average factor of 1/2. If collection-year estimates are being calculated, then the factor is 1/2 for each month. The factor for calculating data-year estimates is given in the fourth column of appendix table H.4. Note that the collection-year may include victimizations occurring in the previous calendar year because of the retrospective nature of the interview.

The domain of interest (i.e., domain) refers to any specific subset of the universe of interest, including the universe of interest itself. If household weights for the entire sample were summed for a 6-month period, an estimate of the number of households in the United States would be generated. Similarly, the sum of the household weights for all of the sample households in urban areas from January through June would be an estimate of the number of households in urban areas.

**Appendix table H.4. NCVS collection-year and data-year weighting factors**

<b>Year</b>	<b>Month</b>	<b>Collection-year factor (CYF)</b>	<b>Data-year factor (DYF)</b>
<i>t</i>	January	1 / 2	0
	February	1 / 2	1 / 12
	March	1 / 2	2 / 12
	April	1 / 2	3 / 12
	May	1 / 2	4 / 12
	June	1 / 2	5 / 12
	July	1 / 2	6 / 12
	August	1 / 2	6 / 12
	September	1 / 2	6 / 12
	October	1 / 2	6 / 12
	November	1 / 2	6 / 12
	December	1 / 2	6 / 12
<i>t + 1</i>	January	0	6 / 12
	February	0	5 / 12
	March	0	4 / 12
	April	0	3 / 12
	May	0	2 / 12
	June	0	1 / 12

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2013.

The implicit assumption behind the factors in appendix table H.4 is that each month contributes equally within collection-year estimates. With data-year estimates, a fraction of the past 6-months contributes to the estimates, where the fraction is proportion to the number of months that can contribute to the estimate.

The estimator for  $Y_{HH}$  (i.e., the total of some variable of interest  $y_i$  in some domain of interest and in a period of time  $t$ ) is the sum of the household weight multiplied by the quantity  $y_i$  for all sample households in the domain of interest in a period of time  $t$ . That is,

$$\hat{Y}_{HHs,t} = \sum_{\substack{\text{sampleHHs} \\ \text{during time } t}} HHWEIGHT_i y_i$$

Note that the statistic  $N$  is a special case of  $Y$ , where  $y_i = 1$  for all units. A distinction exists between the two estimators because the GVF is appropriate for estimators such as  $N$  but is not appropriate for estimators such as  $Y$ .

**Estimators of person-level totals.** The estimator for  $N_{persons}$  (i.e., the total number of persons in a domain of interest) is the sum of the person weight for all sample persons in the domain of interest within a 6-month period. That is,

$$\hat{N}_{persons} = \sum_{\substack{\text{sample} \\ \text{persons} \\ \text{in 6 months}}} PERSONWEIGHT_i$$

and the estimator for  $Y_{persons, t}$  (i.e., the total for some variable of interest  $y_i$  in the domain of interest and in a period of time  $t$ ) is the sum of the person weight multiplied by the quantity  $y_i$  for all sample persons in the domain of interest and in a period of time  $t$ . That is,

$$\hat{Y}_{personst} = \sum_{\substack{\text{sample} \\ \text{persons} \\ \text{duringtime } t}} PERSONWEIGHT_i y_i$$

**Example H1: Calculate the estimated number of females involved in violent crimes in 2009**

To estimate the number of females involved in violent crimes in 2009, first note that the statistic of interest is a total of the number of persons and, therefore, will sum to the person weight. The domain of interest is females who are victims of violent crime. The estimator is the sum of the person weights for all of the persons in the domain of interest collected during 2009. That is,

$$\hat{N}_{\substack{\text{womenwho} \\ \text{arevictims of} \\ \text{violentcrimes} \\ \text{in2009}}} = \frac{1}{2} \sum_{\substack{\text{sample} \\ \text{womenwho} \\ \text{arevictims of} \\ \text{violentcrime} \\ \text{collected in 2009}}} PERSONWEIGHT_i.$$

The estimate is multiplied by 1/2 because weighting a full year of persons will overestimate the population total by a factor of two.

**Estimators of total incidents.** The estimator for  $N_{\substack{\text{person} \\ \text{incidents} \\ \text{duringtime } t}}$  (i.e., the total number of person

victimizations in some domain interest and in a period of time  $t$ ) is the sum of all incident weights for all sample-person victimizations in the domain of interest and in a period of time  $t$ . That is,

$$\hat{N}_{\substack{\text{person} \\ \text{incidents} \\ \text{duringtime } t}} = \sum_{\substack{\text{sample-person} \\ \text{incidents} \\ \text{duringtime } t}} INCIDENTWEIGHT_i$$

and the estimator for  $N_{\substack{\text{HH incidents} \\ \text{duringtime } t}}$  (i.e., the total number of household incidents for some domain of interest and in a period of time  $t$ ), is the sum of all incident weights of all sample-household victimizations in the domain of interest and in a period of time  $t$ . That is,

$$\hat{N}_{\substack{\text{HH incidents} \\ \text{duringtime } t}} = \sum_{\substack{\text{sample-HH} \\ \text{incidents} \\ \text{duringtime } t}} INCIDENTWEIGHT_i.$$

**Estimators of total victimizations.** The estimator for  $N_{\substack{\text{person} \\ \text{victimizations} \\ \text{duringtime } t}}$  (i.e., the total number of person

victimizations for some domain of interest and in a period of time  $t$ ), is the sum of the victimization weights for all sample-person victimizations in the domain of interest and in a period of time  $t$ . That is,

$$\hat{N}_{\substack{\text{person} \\ \text{victimizations} \\ \text{during time } t}} = \sum_{\substack{\text{sample-person} \\ \text{victimizations} \\ \text{during time } t}} VWGT_i$$

and the estimator of  $N_{\substack{HH \\ \text{victimizations} \\ \text{during time } t}}$  (i.e., the total number of household victimizations for some domain of interest and in a period of time  $t$ ), is the sum of victimization weight for all sample-household victimizations in the domain of interest and in a period of time  $t$ . That is,

$$\hat{N}_{\substack{HH \\ \text{victimizations} \\ \text{during time } t}} = \sum_{\substack{\text{sample-HH} \\ \text{victimizations} \\ \text{during time } t}} VWGT_i$$

Note that  $N_{\substack{HH \\ \text{incidents} \\ \text{during time } t}} = N_{\substack{HH \\ \text{victimizations} \\ \text{during time } t}}$ , because only one household can be a victim in a household incident.

**Estimators of the annual incident rates.** The annual person-level incident rate represents the ratio of total number of person incidents that occurred during year  $t$  with the total population and is represented as  $R_{\substack{\text{person} \\ \text{incidents} \\ \text{for year } t}} = N_{\substack{\text{person} \\ \text{incidents} \\ \text{during year } t}} / N_{\substack{\text{persons} \\ \text{in year } t}}$ . As with the other estimators, the incident rate

can be defined in terms of specific domains of interest. The collection-year estimator is defined as—

$$\hat{R}_{\substack{\text{person} \\ \text{incidents} \\ \text{for year } t}} = \frac{\hat{N}_{\substack{\text{person} \\ \text{incidents} \\ \text{collected in year } t}}}{\hat{N}_{\substack{\text{persons} \\ \text{in year } t}}} = \frac{\sum \text{INCIDENTWEIGHT}_i}{\frac{1}{2} \sum_{\substack{\text{sample-persons} \\ \text{collected in year } t}} \text{PERSONWEIGHT}_i}$$

and the data-year estimator is defined as—

$$\begin{aligned} \hat{R}_{\substack{\text{person} \\ \text{incidents} \\ \text{for year } t}} &= \frac{\hat{N}_{\substack{\text{person} \\ \text{incidents} \\ \text{occured in year } t}}}{\hat{N}_{\substack{\text{persons} \\ \text{in year } t}}} \\ &= \frac{\sum_{\substack{\text{sample-person} \\ \text{incidents} \\ \text{occured in year } t}} \text{INCIDENTWEIGHT}_i + \sum_{\substack{\text{sample-person} \\ \text{incidents} \\ \text{occured in year } t+1}} \text{INCIDENTWEIGHT}_i}{\sum_{\substack{\text{sample-persons} \\ \text{in year } t}} \text{DYF}_{t,\text{month}} \times \text{PERSONWEIGHT}_i + \sum_{\substack{\text{sample-persons} \\ \text{in year } t+1}} \text{DYF}_{t+1,\text{month}} \times \text{PERSONWEIGHT}_i} \end{aligned}$$

The household incident rate is calculated the same way as the person incident rate, except the person weights are replaced by the household weights, and sample-household victimizations in the numerator and sample households in the denominator are summed.

Remember that when estimating an annual incident rate, the sum of the denominator should be adjusted by either the collection-year or data-year adjustment factor, otherwise it will be twice as large as it should be. For data-year estimates, sum all of the incidents that were reported in year  $t$ , and for collection-year estimates, sum all of the incidents that were collected in year  $t$ . This applies to all of the estimators of rates in the remainder of this section.

**Estimators of victimization rates.** The household-level victimization rate represents the ratio of total number of households victimized in a period of time  $t$  with the total number of households and is represented as  $R_{HH,t} = N_{HH}^{victimizations\ in\ time\ t} / N_{HHs}$ . As with the other estimators, the victimization

rate can be defined in terms of specific domains of interest. The collection-year estimator is defined as—

$$\hat{R}_{HH}^{victimizations\ for\ year\ t} = \frac{\hat{N}_{HH}^{victimizations\ collected\ in\ time\ t}}{\hat{N}_{HHs}^{in\ year\ t}} = \frac{\sum_{sample-HH\ victimizations\ collected\ in\ time\ t} VWGT_i}{\frac{1}{2} \sum_{sampleHHs\ in\ year\ t} HHWEIGHT_i}$$

and the data-year estimator is defined as—

$$\hat{R}_{HH}^{victimizations\ for\ year\ t} = \frac{\hat{N}_{HH}^{victimizations\ occurred\ in\ year\ t}}{\hat{N}_{HHs}^{in\ year\ t}} = \frac{\sum_{sample-HH\ victimizations\ occurred\ in\ year\ t} VWGT_i + \sum_{sample-HH\ victimizations\ occurred\ in\ year\ t+1} VWGT_i}{\sum_{sampleHHs\ in\ year\ t} DYF_{t,month} \times HHWGT_i + \sum_{sampleHHs\ in\ year\ t+1} DYF_{t+1,month} \times HHWGT_i}$$

The person victimization rate is calculated the same way as the household victimization rate, except the household weights are replaced by the person weights, and sample-person victimizations in the numerator and sample persons in the denominator are summed.

Remember that when estimating an annual victimization rate, the sum of the denominator should be adjusted by either the collection-year or data-year adjustment factor, otherwise it will be twice as large as it should be. For data-year estimates, sum all of the victimizations that were reported in year  $t$ , and for collection-year estimates, sum all of the victimizations that were collected in year  $t$ .

Note that both the household and person rates are not proportions since a given individual (either household or person, depending on the rate) can be counted more than one time in the numerator of the rate. An individual is counted more than one time if they reported more than one incident.

**Example H2: Calculate the estimated victimization rate for burglary in rural households in 2008**

To calculate the estimated burglary rate for rural households in 2008, note that two totals contribute to this ratio. The total of the numerator is the estimated number of burglarized rural households in 2008, and the denominator is the estimated number of all rural households in 2008, which is defined as—

$$\hat{R}_{2008}^{burglaryrate\ for\ rural\ HHs} = \frac{\sum_{\substack{\text{sampleruralHHs} \\ \text{burglarizd} \\ \text{in 2008}}} VWGT_i}{\sum_{\substack{\text{sampleruralHHs} \\ \text{burglarizd} \\ \text{in 2008}}} HHWEIGHT_i}.$$

**Rates are not proportions**

Note that neither the household or person rate is a proportion because any given household or person may be involved in multiple incidents or victimized multiple times. However, rates are sometimes treated as proportions in estimation for reasons of convenience. For example, it may be easier for some to model victimization as a proportion using logistic regression, which has a well-established set of developed software, than to use other models for rates. Generally, treating incident and victimization rates as proportions is methodologically acceptable because victimization rates are generally small.

**How to use Generalized Variance Functions (GVF) to estimate variances**

Although replicate weights have many advantages over GVFs, some variance estimates from replicate weights can be quite variable and computationally demanding. GVFs are easier to use than replicate weights, and they stabilize variance estimates and are computationally more efficient than variances estimated from replicate weights. Wolter (1985) provides a technical introduction to GVFs and gives additional reasons why GVFs are often used. For these reasons, the Census Bureau calculates GVF parameters for the NCVS.

To estimate GVF parameters for the NCVS, the Census Bureau first uses the final weights to estimate totals of the 43 major crime categories. The totals are estimated for all of the persons and households, and are estimated by a variety of key domains. Overall, hundreds of estimates are calculated. The replicate weights are then used to calculate direct variance estimates for all of the totals.

**Why GVFs are used**

GVFs are easy to calculate, and because variance estimates are based on sample data, these estimates have variances of their own. The estimated variance for a survey estimate generally has less precision than the survey estimate itself. This means that the estimates of variance for the same statistic may vary considerably from year-to-year or for related characteristics in a given

year. GVFs provide some stability to the estimates of variance by averaging the variances from estimates of similar size.

### How to calculate the variance with a GVF

Let  $\hat{N}$  be an estimator of a total number of units (i.e., sum of the weights) within a domain interest and let  $v(\hat{N})$  be its variance. The variance of  $v(\hat{N})$  is then calculated as a function of  $\hat{N}$  as—

$$\frac{v(\hat{N})}{\hat{N}^2} = a + \frac{b}{\hat{N}} + \frac{c}{\sqrt{\hat{N}}}$$

where  $a$ ,  $b$ , and  $c$  are the parameters of the model.

The parameters  $a$ ,  $b$ , and  $c$  of expression (1) are provided each year for four different types of estimates. Separate GVFs are provided for property and person crimes. The overall property and person GVFs can be used to produce variances for any of the crime categories at the national level. The domain GVFs can be used to produce variances for subsets of the data based on geographic, demographic, and other characteristics.

#### ***Example H3: Estimating the variance of the estimated number of 2011 households victimized by theft with a GVF***

Let us estimate the variance of the 2011 estimate of the number of households victimized by theft with a GVF. The first step is to get the GVF parameters and the estimate of the statistic of interest (i.e., the number of 2011 households victimized by theft). In 2011, 12,821,090 completed thefts were reported.

Next, the appropriate GVF parameters for 2011 are used. The domain of interest is overall theft. The parameter for overall property crime estimates is used. Appendix table H.5 provides the parameters for the GVF of the 2011 NCVS estimates.

**Appendix table H.5. NCVS GVF parameter set, including series crimes, 2011**

GVF	<i>a</i>	<i>b</i>	<i>c</i>
Overall person crime estimates	-0.00060211	2439	9.511
Person crime domain estimates	-0.00081383	2309	12.916
Overall property crime estimates	-0.00028723	4182	2.809
Property crime domain estimates	-0.00037148	2981	3.852

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2011.

Inserting the estimate of theft and parameters generates the square of the CV:

$$\frac{v(\hat{N})}{\hat{N}^2} = a + \frac{b}{\hat{N}} + \frac{c}{\sqrt{\hat{N}}} = -0.00028723 + \frac{4,182}{12,821,090} + \frac{2.809}{\sqrt{12,821,090}} = (0.029)^2$$

The computation yields a CV of 0.029 and standard error of 367,910. Note that the direct estimate of the CV is 0.028 and the estimate of the standard error is 363,098, which is not the same as the GVF estimate.

## Limitations of GVFs

GVF parameters are estimated from models used to estimate the variance of estimated totals of persons or households. However, GVFs may be less accurate for totals of a characteristic (e.g., the total household income), which were not used in the models.

As with any model, estimates should only be applied to GVFs that are within the range of values used to estimate the parameters.

## How to estimate the variance of a difference at two different times

A statistic that can be used to measure the change in victimization rates for 2 consecutive years is the difference in the two rates (i.e.,  $R_t - R_{t-1}$ ). In this notation, the variable  $t$  is an index on the year of the estimate. The variance for this difference is complicated because the two rates have sample units in common. Therefore, the sample used to estimate each of the rates is not independent. Although having units in common complicates the variance, it is an intentional design feature because it increases the correlation of the estimates of 2 consecutive years.

To estimate the variance of the difference, the covariance between each estimator is accounted for by using the following variance estimator:

$$\hat{v}(\hat{R}_t - \hat{R}_{t-1}) = \hat{v}(\hat{R}_t) + \hat{v}(\hat{R}_{t-1}) - \frac{2\hat{\rho}(\hat{R}_t, \hat{R}_{t-1})}{\sqrt{\hat{v}(\hat{R}_t)}\sqrt{\hat{v}(\hat{R}_{t-1})}}$$

Estimates of the year-to-year corrections, or  $\hat{\rho}(\hat{R}_t, \hat{R}_{t-1})$ , are provided with the GVF parameters  $a$ ,  $b$ , and  $c$ . The estimates of  $\hat{v}(\hat{R}_t)$  and  $\hat{v}(\hat{R}_{t-1})$  should be calculated using the GVF as previously described.

### *Example H4: Calculating the variance of a difference of two rates at two different times*

In this example, estimates of attempted crimes of violence are used as a starting point—18.7 per 1,000 in 2008 and 15.7 per 1,000 in 2009. Using the GVF, the estimated variance of the estimates is calculated as 0.00000171 for 2008 and 0.00000106 for 2009. The last piece needed is the year-to-year correction, which is 0.37. With all of the pieces, the variance of the difference is calculated as—

$$\begin{aligned} \hat{v}(\hat{R}_{t=2009} - \hat{R}_{t=2008}) &= \hat{v}(\hat{R}_{t=2009}) + \hat{v}(\hat{R}_{t=2008}) - \frac{2\hat{\rho}(\hat{R}_{t=2009}, \hat{R}_{t=2008})}{\sqrt{\hat{v}(\hat{R}_{t=2009})}\sqrt{\hat{v}(\hat{R}_{t=2008})}} \\ &= (0.00000171) + (0.00000106) - \frac{2(0.37)}{(0.00131) \times (0.00103)} \\ &= -548,452 \end{aligned}$$

The negative variance indicates that the general correlation should not be applied to this specific case. When the variance of each rate is small, it is typically not necessary to account for a correlation. By substituting a zero for the correlation, the variance calculation would be—

$$\begin{aligned} \hat{v}(\hat{R}_{t=2009} - \hat{R}_{t=2008}) &= (0.00000171) + (0.00000106) - 0 \\ &= 0.0000277 \end{aligned}$$

This estimate is different from the estimate generated by estimating the variances directly with replicate weights. (See also example H 5.)

### How to estimate the variance of a rate with a GVF

In addition to totals, the following examples describe how to calculate the estimated variance for two important statistics: the rate and the difference between two rates.

**Estimator of the variance for an estimated victimization rate.** Assuming that the estimator of the victimization rate is of the form—

$$\hat{R}_{HH} = \frac{\hat{N}_{HH-level\ victimizations}}{\hat{N}_{HHs}} = \frac{\sum_{sample\ HH\ victimizations} VWGT_i}{\sum_{sample\ HHs} HHWEIGHT_i}$$

where  $\hat{N}_{HH-level\ victimizations}$  is the estimator of the number of household victimizations and  $\hat{N}_{HHs}$  is the estimator of the total number of units, an estimator of the variance for a victimization rate is

$$\hat{v}(\hat{R}_t) = \frac{b\hat{R}(1-\hat{R})}{\hat{N}_{HHs}} + \frac{c\hat{R}(\sqrt{\hat{R}}-\hat{R})}{\sqrt{\hat{N}_{HHs}}}. \text{ The variance for the estimator of the person-level}$$

victimization rate is found similarly.

#### **Example H 5: Calculating the variance of a rate with GVF parameters**

In this example, the variance of the 2010 theft rate in the northeast region is calculated. The theft rate in this region was 73.7 per 1,000. Because a geographic subset of the full sample is being examined, the correct GVF parameter for the property crime domain is the correct choice. As shown in Appendix D, the *b* parameter is 3,297 and the *c* parameter is 1.687. The weighted number of total households in the northeast region is 22,452,095. The result of putting all of these pieces into the above formula is—

$$\begin{aligned} \hat{v}(\hat{R}_t) &= \frac{3,297 \times 0.0737(1-0.0737)}{22,452,095} + \frac{1.687 \times 0.0737(\sqrt{0.0737}-0.0737)}{\sqrt{22,452,095}} \\ &= 0.0000152 \end{aligned}$$

For comparison, if the variance using replicate weights is computed directly, the variance estimate is 0.0000169.

### How to calculate variance estimates using replicate weights

Starting with the 2011 estimates, replicate weights are provided on the public use files. Data users will be able to estimate variances directly using specialized statistical software. The four sets of replicate weights (i.e., household, person, incident, and victimization) will have 160 sets of weights that will produce 160 different replicates.

The replicate weights can be used to estimate any general statistic  $\theta$ , where  $\hat{\theta}$  is either a linear or a nonlinear function of totals  $Y$ . The variance estimator is generally stated as—

$$\hat{v}(\hat{\theta}) = \frac{4}{R} \sum_{r=1}^R (\hat{\theta} - \hat{\theta}_r)^2 \quad (1)$$

where the estimator of the replicate total for replicate  $r$  is defined as a function of linear estimates of a total  $y_{ij}$  as—

$$\hat{\theta}_r = \theta \left( \sum_{i=1}^n w_{ir} y_{i,j=1}, \sum_{i=1}^n w_{ir} y_{i,j=2}, \dots, \sum_{i=1}^n w_{ir} y_{i,j=p} \right) \quad (2)$$

where

$i$  = index on units in the sample

$j$  = index on the characteristic of interest

$n$  = the sample size

$y_{ij}$  = the variable of interest for unit  $i$  and characteristic  $j$

$w_{ir}$  = the replicate weight for unit  $i$  in replicate  $r$

and the estimator of the mean of the replicate totals is—

$$\hat{\theta} = \frac{1}{R} \sum_{r=1}^R \hat{\theta}_r \quad (3)$$

Note that the factor of 4 is necessary because the method used by Fay and Train (1995) is used with replicate factors of 0.5 and 1.5.

***Example H 6: Using replicate weights to estimate the variance***

This example shows the basic steps of calculating a variance with replicate variances, for data users who are not familiar with replicate weights. Suppose the following sample of  $n = 5$  household units in appendix table H.6 is observed:

**Appendix table H.6. Example of using NCVS replicate weights**

Sample household unit	Motor vehicle theft	Sample weight	Replicate weight			
			Replicate 1	Replicate 2	Replicate 3	Replicate 4
Unit #1	Yes	<b>15.96</b>	<b>15.96</b>	<b>5.30</b>	<b>24.90</b>	<b>15.84</b>
Unit #2	No	24.47	24.47	46.06	22.46	7.29
Unit #3	Yes	<b>20.21</b>	<b>20.21</b>	<b>22.38</b>	<b>5.57</b>	<b>34.11</b>
Unit #4	Yes	<b>17.02</b>	<b>17.02</b>	<b>18.85</b>	<b>26.56</b>	<b>5.07</b>
Unit #5	No	22.34	22.34	7.42	20.51	37.70
<b>Total</b>		<b>53.19</b>	<b>53.19</b>	<b>46.53</b>	<b>57.03</b>	<b>55.02</b>

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2013.

And the statistic of interest to be estimated is the total number of motor vehicle thefts. The full sample estimate is the sum of the weights for all of the sample units with a motor vehicle theft.

$$\text{Full sample estimate } \hat{\theta} = 15.96 + 20.21 + 17.02 = 53.19$$

Now using replicate weights,  $\theta$  is estimated for each replicate using the replicate weights.

$$\text{Replicate 1 estimate } \hat{\theta}_{r=1} = 15.96 + 20.21 + 17.02 = 53.19$$

$$\text{Replicate 2 estimate } \hat{\theta}_{r=2} = 5.30 + 22.38 + 18.85 = 46.53$$

$$\text{Replicate 3 estimate } \hat{\theta}_{r=3} = 24.90 + 5.57 + 26.56 = 57.03$$

$$\text{Replicate 4 estimate } \hat{\theta}_{r=4} = 15.84 + 34.11 + 5.07 = 55.02$$

Using (1), where  $R = 4$  (i.e., the number of replicates from the example), the variance estimate is calculated for the total number of motor vehicle thefts as—

$$\begin{aligned} \text{var}(\hat{\theta}) &= \frac{4}{4} [(53.19 - 53.19)^2 + (46.53 - 53.19)^2 + (57.03 - 53.19)^2 + (55.02 - 53.19)^2] \\ &= 62.45 \end{aligned}$$

The survey estimate of the total number of motor vehicle thefts is 53.19 with an estimated variance of 62.45, or a standard error of 7.90.

***Example H 7: Calculating the variance of the 2011 female assault victimization rate with replicate weights***

In this example, the variance of the estimated 2011 assault rate for female victimizations is calculated. The example illustrates an important point of the replicate weights. That is, how the different sets of replicate weights should be used with estimators that include different types of weights.

This example illustrates calculating variances using replicate weights. The source of this example does not take series crimes into account. Therefore, some raw numbers may differ from

other estimates that include the series crimes. The process in this example is the same, regardless of whether or not series crimes are included in the crime rates.

In the representation of replications in equation (2), only one weight is used. However, NCVS has four different replicate weights that should be used together to make replicate estimates. For this example, a victimization rate is computed. Therefore, the victimization weight will be used in the numerator, and the person weight for the year will be used in the denominator. Because estimates are computed for an entire year, the original person weight must be multiplied by a factor of 1/2 to compute the person weight for the year. See appendix table H.6 for more information. For the first replicate, the first replicate victimization weight and the first replicate person weight for the year would be used.

The formula for the replicate estimate for this example would be—

$$\hat{\theta}_r = \frac{\sum_{\substack{\text{person} \\ \text{victimizations involving} \\ \text{assault and} \\ \text{females}}} VWEIGHT_{ir}}{\sum_{\substack{\text{persons that} \\ \text{are females}}} PERSONWEIGHT_{ir}}$$

Appendix table H 7 summarizes the weights that are needed to calculate the overall estimate. The first column of appendix table H 7 identifies all of the females in the sample. Female #3 was assaulted twice. Therefore, both victimizations need to be counted in the estimates. Two records exist for this person. The second column indicates whether or not a person was a victim of an assault. The third column is the person weight, and the fourth column is the victimization weight.

**Appendix table H 7. Weights needed for replicate weight calculations for 2011 female assault victimization rate**

Sample female	Victim of assault	Person weight	Victimization weight
Female #1	No	1,414.66	0
Female #2	Yes	1,371.00	1,502.77
Female #3	Yes	1,243.21	2,486.42
Female #3	Yes	1,243.21	2,486.42
Female #4	No	1,534.71	0
...	...	...	...
Female last	No	1,948.15	0
Overall total		131,521,046	1,670,227
Overall rate*		$\hat{\theta} = 12.7$	

\*Overall rate per 1,000 females age 12 or older.

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2011.

Appendix table H 8 illustrates the process for calculating replicate estimates using replicate weights. Replicate estimates are calculated in the same manner as the overall estimate, except

that replicate weights are used. Again, because female #3 was assaulted twice in the year, two records exist for this person.

**Appendix table H 8. Replicate weight calculations for 2011 female assault victimization rate**

Sample person	Replicate weight						
	Replicate 1 numerator	Replicate 1 denominator	Replicate 2 numerator	Replicate 2 denominator	...	Replicate 160 numerator	Replicate 160 denominator
Female #1	1,382.18	0	1,399.96	0		411.120	0
Female #2	2,382.45	2,611.42	670.80	735.27		651.65	714.28
Female #3	1,227.35	2,454.70	1,202.73	2,405.47		1,239.41	2,478.82
Female #3	1,227.35	2,454.70	1,202.73	2,405.47		1,239.41	2,478.82
Female #4	1,515.28	0	1,529.88	0		439.73	0
...	...	...	...	...		...	...
Female last	1,936.56	0	1,964.03	0		570.78	0
Overall total	1,626,905	131,502,567	1,697,363	131,510,305	...	1,626,905	131,502,567
Overall rate*	$\hat{\theta}_{r=1} = 12.4$		$\hat{\theta}_{r=2} = 12.9$		...	$\hat{\theta}_{r=160} = 13.3$	

\*Overall rate per 1,000 females age 12 or older.

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2011.

Using (1), where  $R = 160$  replicates, the variance estimate for the total number of female assault victimizations is calculated as—

$$\hat{v}(\hat{\theta}) = \frac{4}{160} [(0.0124 - 0.0127)^2 + (0.0129 - 0.0127)^2 + \dots + (0.0133 - 0.0127)^2]$$

$$= 0.00000058$$

## References

- American Association for Public Opinion Research. (2011). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 7<sup>th</sup> edition. AAPOR.
- Biderman, A.D., & Cantor, D. (1984). A Longitudinal Analysis of Bounding, Respondent Conditioning and Mobility as Sources of Panel Bias in the National Crime Survey. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*.
- Brick, J.M., & Kalton, G. (1996). Handling Missing Data in Survey Research. *Statistical Methods in Medical Research*, 5, 215–238.
- Bureau of Justice Statistics. (2012). “Terms & Definitions: Victims.” Retrieved from <http://www.bjs.gov/index.cfm?ty=tdtp&tid=9>
- Cochran, W.G. (1977). *Sampling Techniques*. New York: John Wiley & Sons.
- Deville, J.C., & Särndal, C.-E. (1992). Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association*, 87(418), 376–382.
- Fay, R.E., & Train, G.F. (1995). Aspects of Survey and Model-Based Postcensal Estimation of Income and Poverty Characteristics for States and Counties. In *Proceedings of the Section on Government Statistics, American Statistical Association*, 154–159.
- Hájek, J. (1981). *Sampling from a Finite Population*. New York: Marcel Dekker, Inc.
- Judkins, D.R. (1990). Fay’s Method for Variance Estimation. *Journal of Official Statistics*, 6(3), 223–239.
- Kish, L. (1965). *Survey Sampling*. New York: John Wiley & Sons.
- Krenzke, T. (1995). Reevaluating Generalized Variance Model Parameters for the National Crime Victimization Survey. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 327–332.
- McCarthy, P.J. (1966). Pseudo-replication: Half-samples. *Review of the International Statistical Institute*, 37, 239–264.
- Neter, J., & Waksberg, J. (1964). A Study of Response Errors in Expenditures Data from Household Interviews. *Journal of the American Statistical Association*, 59, 17–55.
- Särndal, C.-E., & Lundström, S. (2005). *Estimation in Surveys with Nonresponse*. Springer-Verlag.
- U.S. Census Bureau. (2006). *Current Population Survey Design and Methodology: Technical Paper 66*.
- Wolter, K.M. (1985). *Introduction to Variance Estimation*. Springer-Verlag.