National Crime Victimization Survey, 2016

Technical Documentation

Contents

Chapter 1. Background

NCVS overview

The 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, and the only source of data including victimizations both reported and not reported to police. The United States Census Bureau (Census Bureau) serves as the primary data collection organization for the NCVS by conducting interviews and processing sample data on a monthly basis.

The NCVS is a self-report survey in which interviewed persons 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 7 interviews over a 3-year period. 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.

The target population of the NCVS is U.S. residents age 12 or older residing in housing units or group quarters (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.

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 the 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, other population subgroups, and the top 22 most populous states[1].

The NCVS collects information on both personal crimes—violent or nonviolent crimes committed against persons—and property crimes, which do not involve direct contact between the offender and victim. 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 is collected for each victimization incident. Incidents reported in the NCVS are associated with the address of the respondent and not with where the incident occurred. Periodically, the survey includes separate supplements on additional topics, such as identity theft, crime in schools, stalking, and contacts between law enforcement and the public.

---

[1] The 22 states identified for state-level estimates are Arizona, California, Colorado, Florida, Georgia, Illinois, Indiana, Maryland, Massachusetts, Michigan, Minnesota, Missouri, New Jersey, New York, North Carolina, Ohio, Pennsylvania, Tennessee, Texas, Virginia, Washington, and Wisconsin.

The BJS maintains a webpage dedicated to the NCVS on its website: https://www.bjs.gov.

Document scope

The purpose of this Technical Documentation is to provide the NCVS data users with information about the history and scope of the NCVS, and the procedures through which NCVS data are collected and processed. This document provides details about the NCVS sampling, data collection, data processing, and weighting and estimation methodologies that remain generally stable unless there are significant sample or survey design changes.

This document serves as a complement to the annually-published NCVS Source and Accuracy Statement which provides statistical information about the data quality of NCVS estimates. Beginning in 2016, the Source and Accuracy Statement is published as part of the Codebook for the NCVS public use files. The most recent Codebook is located here: https://www.icpsr.umich.edu/icpsrweb/ICPSR/series/95.

Additionally, the information in this document is based on the 2010 NCVS sample design, which was phased into production with 2015 data collection. Users seeking sampling information based on the 2000 sample design should refer to the 2013 version of the NCVS Technical Documentation, which is located here: https://www.bjs.gov/content/pub/pdf/ncvstd13.pdf.

Survey history

In 1967, the President's Commission on Law Enforcement and Administration of Justice issued a report concluding that more data were needed on crimes not reported to the police. The commission recommended the development of a nationwide crime victimization survey to fill that need.

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 Federal Bureau of Investigation (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 the "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 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, or 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 1992 redesign may be found in the following documents:
https://www.bjs.gov/content/pub/pdf/ERVE.PDF
https://www.bjs.gov/content/pub/pdf/ncsrqa.pdf

In July 2006, the NCVS converted to a fully automated Computer-Assisted Personal Interviewing (CAPI) environment. Several additional methodological improvements have been researched in recent years. Data users can find information on these enhancements on the NCVS Research and Development page located on the BJS website, here: https://www.bjs.gov/index.cfm?ty=tp&tid=91.

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 death, 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 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.

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 classification

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

Table 1. NCVS crime classification

Personal crimes	Property crimes
Crimes of Violence	Burglary
Completed	Completed
Attempted	Forcible entry
Serious violent crimes*	Unlawful entry without force
Rape/sexual assault	Attempted forcible entry
Rape	Motor vehicle theft
Completed	Completed
Attempted	Attempted
Sexual assault	Theft
Robbery	Completed
Completed	Less than $50
With injury	$50-$249
Without injury	$250 or more
Attempted	Amount not available
With injury	Attempted
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	

* Serious violent crimes include rape/sexual assault, robbery, and aggravated assault.

The key measures for the NCVS are rates and totals for:

- crimes of violence
- serious violence (rape/sexual assault, robbery, and aggravated assault)
- property crimes
- crimes not reported to the police

Key domains are:

- sex
- race/Hispanic origin
- age

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 crime 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. 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.

The core NCVS does not measure some types of crimes, including homicide, kidnapping, verbal threats over the phone, and other forms of crime involving social media, arson, 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 possession or sales, prostitution, sex trafficking, illegal gambling, con games, and blackmail are also not measured.

Rape and sexual assault estimates

The NCVS uses a two-stage measurement approach in the screening and classification of criminal victimization, including rape and sexual assault. In the first stage of screening, survey respondents are administered a series of "short-cue" screening questions designed to help respondents think about different experiences they may have had during the reference period (see NCVS-1 at https://www.bjs.gov).

This design improves respondent recall of events, particularly for incidents that may not immediately come to mind as crimes, such as those committed by family members and acquaintances. Respondents who answer affirmatively to any of the short-cue screening items are subsequently administered a crime incident report (CIR) designed to classify incidents into specific crime types (see NCVS-2 at https://www.bjs.gov).

First stage of measurement. Two short-cue screening questions are specifically designed to target sexual violence:

1. Other than any incidents already mentioned, has anyone attacked or threatened you in any of these ways—

(a) with any weapon, such as a gun or knife

(b) with anything like a baseball bat, frying pan, scissors, or stick

(c) by something thrown, such as a rock or bottle

(d) include any grabbing, punching, or choking

(e) any rape, attempted rape, or other types of sexual attack

(f) any face-to-face threats.

OR

(g) any attack or threat or use of force by anyone at all? Please mention it even if you are not certain it was a crime.

2. Incidents involving forced or unwanted sexual acts are often difficult to talk about. Other than any incidents already mentioned, have you been forced or coerced to engage in unwanted sexual activity by—

(a) someone you did not know

(b) a casual acquaintance

OR

(c) someone you know well?

Respondents may screen into a CIR if they respond affirmatively to another short-cue screening question. For instance, a separate screening question cues respondents to think of attacks or threats that took place in specific locations, such as at home, work, or school. A respondent who recalled a sexual victimization that occurred at home, work, or school and answered affirmatively would be administered a CIR even if they did not respond affirmatively to the screening question targeting sexual violence.

Second stage of measurement. The CIR is used to collect information on the attributes of each incident. The key attributes of sexual violence that are used to classify a victimization as a rape or sexual assault are the type of attack and physical injury suffered. Victims are asked if "the offender hit you, knock[ed] you down, or actually attack[ed] you in any way;" if "the offender TR[IED] to attack you;" or if "the offender THREATEN[ED] you with harm in any way?" The survey participant is classified as a victim of rape or sexual assault if he or she responds affirmatively to one of these three questions and then responds that the completed, attempted, or threatened attack was—

• rape

• attempted rape

• sexual assault other than rape or attempted rape

• verbal threat of rape

• verbal threat of sexual assault other than rape

• unwanted sexual contact with force (e.g., grabbing, fondling)

• unwanted sexual contact without force (e.g., grabbing, fondling).

Whether the victim selects one of these response options to describe the attack, he or she is also classified as a victim of rape or sexual assault if the injuries suffered as a result of the incident are described as:

• rape

• attempted rape

• sexual assault other than rape or attempted rape

Classification of rape and sexual assault in the National Crime Victimization Survey

Measures	Elements of sexual violence
Completed rape	Type of attack = rape
	Type of injury = rape
Attempted rape	Type of attack = attempted rape
	Type of injury = attempted rape
	Type of threat = verbal threat of rape with weapon
Sexual assault	Type of attack = sexual assault other than rape or attempted rape
	Type of injury = sexual assault other than rape or attempted rape
	How try attack/threaten = unwanted sexual contact with or without force

	How try attack/threaten = verbal threat of sexual assault other than rape

Note: Victim is determined to be present in all measures of rape and sexual assault. Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2016.

Coercion. Although the CIR does not ask respondents if psychological coercion was used, one screening question targeted to rape and sexual violence asks respondents if force or coercion was used to initiate unwanted sexual activity.

The final classification of incidents by the CIR results in the following definitions of rape and sexual assault used in the NCVS:

Rape. Coerced or forced sexual intercourse. Forced sexual intercourse means vaginal, anal, or oral penetration by the offender(s). This category could include incidents where the penetration was from a foreign object such as a bottle. Includes attempted rapes, male and female victims, and both heterosexual and same-sex rape. Attempted rape includes verbal threats of rape.

Sexual assault. A wide range of victimizations, separate from rape or attempted rape. These crimes include attacks or attempted attacks generally involving unwanted sexual contact between victim and offender. Sexual assaults may or may not involve force and include such things as grabbing or fondling. Sexual assault also includes verbal threats.

Supplements

Additional sets of questions, or "supplements," are regularly added to the end of the NCVS interview to make timely estimates of specific types of victimization. Typically, a supplement is in the field for 6 months (January-June or July-December), after which it is replaced by a different supplement with a new topic. See Appendix C for a full list of the NCVS supplements by topic and date fielded. Each supplement has its own Source and Accuracy Statement. Users seeking detailed technical information about a particular supplement can find it in the Source and Accuracy Statement for that supplement.

Listed below is general information about NCVS supplements fielded in recent years.

Identity Theft Supplement (ITS)

The purpose of the ITS is to measure the prevalence of identity theft among persons, the characteristics of identity theft victims, and patterns of reporting to the police, credit bureaus, and other authorities. The ITS was also designed to collect important characteristics of identity theft such as how the victim's personal information was obtained; the physical, emotional and financial impact on victims; offender information; and the measures people take to avoid or minimize their risk of becoming an identity theft victim. Respondents are asked to report if they have experienced the misuse of an existing credit card, the misuse of another existing account (banking, savings, etc.), the misuse of personal information to open a new account, or the misuse of personal information for other fraudulent purposes in the 12 months prior to the interview. All self-interviewed NCVS household members age 16 and older are eligible for the ITS.

Police Public Contact Survey (PPCS)

The PPCS provides estimates on the nature and consequences of respondents' interactions with police, and the possible use of excessive force by law enforcement officers during these interactions. Screener questions first ascertain whether respondents have had any contacts with the police during the past 12 months. For respondents who report contact with the police in one more screener questions, a series of questions collects, for the most recent interaction, information on the reason for the police contact, the demographic characteristics of the police officer(s), respondent actions toward the police officer and vice versa (including threats and physical violence), and the outcome of the interaction (including tickets and arrests). All self-interviewed NCVS household members age 16 and older are eligible for the PPCS.

School Crime Supplement (SCS)

Cosponsored by the National Center for Education Statistics (NCES) and the BJS, the SCS is designed to gather additional information about school-related victimizations to complement basic information collected in the NCVS. The supplement asks a number of key questions about crime and violence inside schools, on school grounds, or on the way to and from school. Specifically, the supplement contains questions on preventative measures employed by the school to deter crime; students' participation in extracurricular activities; transportation to and from school; students' perception of rules and equality in school; bullying and hate crime in school; the presence of street gangs in school; availability of drugs and alcohol in the school; attitudinal questions relating to the fear of victimization in school; access to firearms; and student characteristics such as grades received in school and postgraduate plans. The SCS is administered to all interviewed NCVS household members who are 12 to 18 years of age, who were enrolled in primary or secondary education programs leading to a high school diploma, and who were enrolled in school sometime during the six months prior to the interview.

Supplemental Fraud Survey (SFS)

The Supplemental Fraud Survey (SFS) is designed to measure the prevalence, characteristics, and consequences of being a victim of personal financial fraud. For the SFS, fraud is defined as intentionally and knowingly deceiving the victim by misrepresenting, concealing, or omitting facts about promised goods, services, or other benefits and consequences that are nonexistent, unnecessary, never intended to be provided, or deliberately distorted for the purpose of monetary gain. The victim must lose money in the transaction in order for the incident to be considered fraud. The supplement collects information on the type of fraud committed, how much money obtained by the offender, how much money was recovered, if the incident was reported to law enforcement or a consumer protection agency, and any negative social, emotional, or financial consequences of the fraud incident. The Census Bureau first collected SFS data from October to December of 2017. All self-interviewed NCVS household members age 18 and older are eligible for the SFS.

Supplemental Victimization Survey (SVS)

The goal of the SVS is to measure the prevalence of stalking victimization among persons, the characteristics of stalking victims, and patterns of reporting to the police and other authorities. The SVS was also designed to collect important characteristics of stalking such as the physical and emotional impact on victims, offender information, measures of self-protective actions, and the criminal justice system response. The survey begins with a screener instrument asking about each

element of stalking as defined by the Violence Against Women Act (2013 as amended). Respondents are classified as stalking victims if they experienced repeated unwanted contacts or behaviors that caused them substantial emotional distress, to fear for their safety or the safety of others, or that would have caused a reasonable person to fear for their safety or the safety of someone they know. If a respondent screens in as a stalking victim based on responses to the screener, the survey continues with the incident portion, which focuses on details of the stalking victimization. All self-interviewed NCVS household members age 16 and older are eligible for the SVS.

Chapter 2. Sample design

Introduction

The NCVS is a panel, or longitudinal, survey in which persons in each housing unit or GQ are interviewed every six months, for a total of seven interviews. The NCVS interviews are conducted continuously throughout the year in a design that divides the NCVS sample into seven rotating groups. Within each of the seven rotating groups, six panels are designated, each of which is interviewed in a given month and every six months thereafter, 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.

The rotation chart in Figure 1 illustrates the sampling and interviewing schedule for eighteen months of NCVS interviewing. The years along the top indicate when the sample was selected, and the years and months along the left side are when the sample cases are interviewed. The two-digit numbers represent the panel and rotation numbers assigned to each sample unit at the time of sampling. In July through December 2019, for example, 2015 rotation 5 cases will be removed from the sample and 2018 rotation 5 cases will be added to the sample. Sample cases in panel 3 within 2018 rotation 5 (the two-digit number 35 in the 2018 column) will begin interviewing in September 2019, will be interviewed again in March 2020, and will interview every six months until their seventh and final interviews September 2022 (not pictured).

Interview Year and Month		Sampling Year				
		2015	2016	2017	2018	2019
2019	JAN	15 16	17 11	12 13	14	
	FEB	25 26	27 21	22 23	24	
	MAR	35 36	37 31	32 33	34	
	APR	45 46	47 41	42 43	44	
	MAY	55 56	57 51	52 53	54	
	JUN	65 66	67 61	62 63	64	
	JUL	16	17 11	12 13	14 15	
	AUG	26	27 21	22 23	24 25	
	SEP	36	37 31	32 33	34 35	
	OCT	46	47 41	42 43	44 45	
	NOV	56	57 51	52 53	54 55	
	DEC	66	67 61	62 63	64 65	
2020	JAN		17 11	12 13	14 15	16
	FEB		27 21	22 23	24 25	26
	MAR		37 31	32 33	34 35	36
	APR		47 41	42 43	44 45	46
	MAY		57 51	52 53	54 55	56
	JUN		67 61	62 63	64 65	66

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2016.

The NCVS sample is a two-stage stratified sample of housing units and GQs within 542 sample areas. The first stage of sampling involves the definition, stratification, and selection of primary sampling units (PSUs), which are defined as a large metropolitan area, county, or group of bordering counties. First-stage sampling occurs once every ten years, which reduces survey costs and allows for consistent data collection from trained interviewers within each PSU area. The 2010 design sample PSUs were sampled using population data from the 2010 census and began interviewing in 2016. The 2000 design sample PSUs were sampled using population data from the 2000 census and interviewed in years 2006 through 2015.

The second stage of sampling selects housing units and GQs within selected PSUs. This within-PSU selection occurs every year for housing units and every three years for GQs. At the time of sampling, all selected units are divided into the rotations and panels that will begin interviewing in the following one or three years. The housing unit and GQ samples are both divided into two rotations per year and six panels (or months) per rotation.

The NCVS has a panel design to allow for analysis of year-to-year changes and to reduce response errors during data collection. 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. The repeated interviews also allow for bounding later interviews to prevent telescoping, or bringing incidents outside the reference period into the period. Bounding is a process to ensure that previously reported incidents are not reported again. It provides a more accurate measure of criminal victimization within the NCVS sample housing unit or GQ.

A bounding weighting adjustment is applied to compensate for the use of the first interview and the potential inclusion of victimizations that are out of the reference period. New construction cases and replacement households entering the sample during the second through seventh interview are not bounded, and do not receive the bounding weighting adjustment.

Sampling frame

A sampling frame is a list of units in the population of interest from which a sample may be drawn. The NCVS selects its sample from two dynamic sampling frames, one for housing units and one for GQs. Both frames are based upon the Master Address File (MAF), which is a national inventory of addresses. The MAF is continually updated by various Census Bureau programs and external sources. Depending on the source of the data, a MAF record may, but does not always, include entries for: mailing address, location description, latitude/longitude coordinates, and census tract/block.

New housing units are added to the MAF, and therefore the NCVS sampling frame, through semiannual updates from a variety of address sources, including the U.S. Postal Service Delivery Sequence File, local government files, and field listing operations. Some external sources do not provide census tract/block information, and all such MAF records are referred to as "ungeocoded" units. Ungeocoded units lack all census geography (such as place, county subdivision, urbanicity, urban area, and principal city status) that is linked to the census block. The NCVS imputes some of this geography using the state, county, and ZIP codes that are available for all ungeocoded units.

Housing unit frame

The housing unit frame is updated twice each year with the January and July MAF data. This process updates address and geography information for existing frame records and adds new construction records to the frame. Included in this process is a MAF filtering step, which designates each frame record as either "valid" (passed the filter and eligible for the housing unit frame) or "invalid" (failed the filter, ineligible for the frame). All MAF records (valid and invalid) are included in the housing unit frame so they have a chance of sample selection in case their status later becomes valid. This unit status is the first sort variable during sampling, so the sample of valid units is not randomly altered by invalid units. The housing unit frame contains information down to the housing unit level, including individual units within apartment complexes. A housing unit is a group of rooms or a single room occupied as separate living quarters or intended for occupancy as separate living quarters. A housing unit may be occupied by a family or one person, as well as by two or more unrelated persons who share the living quarters.

GQ frame

The GQ frame for the NCVS and the other Census-administered current surveys is created once every three years. The GQ frame consists only of GQs that are, or may potentially be, in-scope for current surveys. Noninstitutional GQs are generally in-scope and are facilities for people who are not under formally authorized and supervised care and custody such as college housing, group homes for adults, workers' living quarters, and convents.[2] Institutional GQs are all out-of-scope and include facilities such as prisons, skilled nursing facilities, and residential treatment centers for

---

[2] Noninstitutional GQs that are out-of-scope include military quarters, military ships, soup kitchens, domestic violence shelters, regularly scheduled mobile food vans, targeted nonsheltered outdoor locations, maritime/merchant vessels, and living quarters for victims of natural disasters.

juveniles. The GQ frame includes some specific types of institutional GQs (such as psychiatric hospitals and non-correctional group homes for juveniles) because they have relatively high rates of converting to in-scope noninstitutional GQs. The GQ frame contains the expected number of units within each GQ, and any GQs selected for sample are sent to the field for a listing of the units (which can be rooms, beds, or people) at the GQ.

Coverage

Some level of overcoverage and undercoverage is expected in any survey frame. Overcoverage occurs when units are duplicated in the frame or when the frame contains units that do not exist (such as previously demolished housing units). Undercoverage occurs when existing units are missing from the frame. The Census Bureau regularly evaluates the coverage of the MAF and implications for the various survey frames. In addition, the NCVS reports coverage rates of survey respondents in the annual Source and Accuracy Statement.

Stratification

The first step of the 2010 design sample involved dividing the United States into first-stage sample units, or PSUs, consisting of large metropolitan areas, counties, or groups of bordering counties. The PSUs were defined within each state using data from the 2010 Census on characteristics such as total land area, current and projected population counts, large metropolitan areas, and natural barriers, such as rivers and mountains. With the exception of some PSUs in Alaska, the required population size for a PSU was at least 7,500 persons. The preferred maximum land area was 3,000 square miles, but some PSUs exceeded this threshold because counties had a large geographic area and small population. These limits were chosen to provide sufficient workload for at least one interviewer per PSU while keeping travel time for individual interviewers to a reasonable limit.

The next step was dividing the PSUs into sampling strata. All PSUs within large Core Based Statistical Areas (CBSAs) were included in the sample with certainty. They are known as self-representing (SR) PSUs because they represent themselves only. All SR PSUs make up their own sampling strata.

The remaining PSUs were non-self-representing (NSR) and were grouped within states with similar NSR PSUs based on decennial census demographic data, American Community Survey data, and administrative crime data. This grouping was an iterative process that randomly grouped PSUs and finally selected the grouping that minimized the between-PSU variance within all strata within the state. Each NSR stratum must also fit the size criteria described above for SR strata regarding interviewer workload and travel time.

First-stage sample selection

The first stage selection identified a sample of PSUs, and this occurs once every ten years. The 2010 design PSUs began interviewing in 2016 and will continue interviewing through 2025. The SR PSUs were automatically selected for sample, and the NSR PSUs were sampled with probability proportional to the population size. One PSU was selected from each NSR stratum.

During the 2000 redesign, the PSU selection algorithm was designed to maximize overlap between the 1990 design and the 2000 design. This minimized the costs and potential impacts to the estimates that are inherent in the staffing and data collection within new areas. Due to the redesign, there was no effort to maximize the amount of overlap between 2010 and 2000 design PSUs in the 2010 design. Plans for future redesigns including using Ohlsson's method of maximizing overlap (Ohlsson, 2000), and the 2010 design provides the first sample which should be completely independent of other samples.

The 2010 design NCVS sample includes 339 SR PSUs and 203 NSR PSUs. These 542 PSUs include 933 counties and county equivalents, 505 of which are "continuing" areas because they were also in sample in the 2000 design NCVS. The remaining 428 counties and county equivalents are new to the NCVS beginning with the 2016 data collection year. Because PSUs were redefined for the 2010 design, some sampled PSUs contain both new and continuing counties. Therefore, the concept of new or continuing first-stage sample areas must be applied at the county level and not at the PSU level.

Second-stage sample selection

For second stage selection, a sample of housing units and GQs is selected from within the first stage sample PSUs. The target sample size, discussed in more detail in the next section, is the number of housing units and GQs expected to result in enough completed person interviews to meet NCVS data reliability goals.

A systematic sample is selected from an ordered list of addresses from the GQ and housing unit frames. As explained by Cochran (1977, p. 208), a systematic 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 NCVS sort variables are listed in the following subsections.

The systematic sample design starts with a known sampling interval referred to as the "take-every" (TE), which is the inverse of the selection probability within a PSU. Next, a random start (RS) is calculated as a random number from a uniform distribution on the interval (0,TE]. 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 are hits and the set of selected units are referred to as the hit string.

Figure 2. Representation of systematic sampling
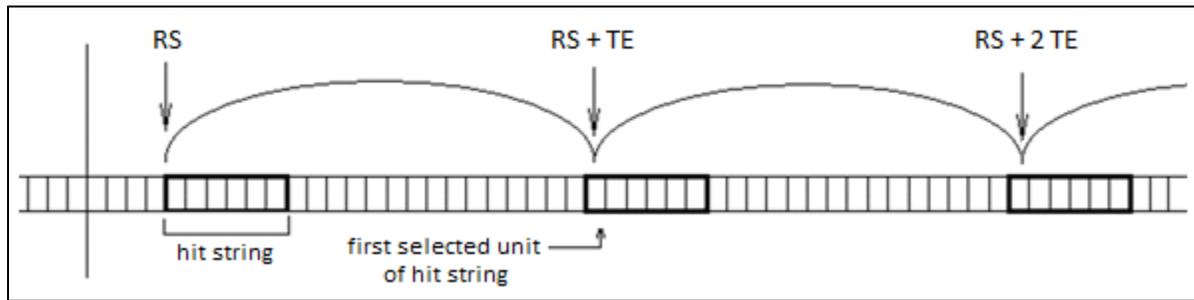
Source: Bureau of Justice Statistics, National Crime Victimization Survey.

Figure 2 illustrates the systematic 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.

The sampling methodology for the second-stage differs slightly by frame, so each frame sample selection is described in more detail below.

Housing unit frame

The housing unit frame second-stage sample selection occurs once a year and selects sample addresses to begin interviewing in the following year. After sample selection, each annual housing unit sample is divided into two rotations so half the sample will begin interviewing in January through June and the other half will begin interviewing in July through December.

The systematic sample design sorts the housing unit frame with respect to the following variables:

- MAF valid/invalid flag
- central city/balance/urban/rural status
- state and county code
- within-PSU stratification code (based on household tenure and tract/block median income)
- tract and block code
- housing unit frame random number
- MAF identification number.

The RS and TE of the systematic sample change each year, but the hit string remains constant at four housing units per sample hit. All housing units within a hit string are assigned to the same interviewing rotation group and month. This is done to reduce field costs by interviewing housing units within close proximity of each other during the same months.

The TE is designed to select twice as much sample as needed to produce the NCVS estimates, and half of the resulting sample is set aside as reserve sample. This reserve sample will not go out for interviews unless special situations, such as separate questionnaire experiments, require additional sample.

Each annual TE is determined based on the size of the housing unit frame within each PSU at the time of sampling. The sampling weights are assigned using this TE so all sampled housing unit weights sum to the total housing units in the universe. To ensure sampling weights always sum to

current population totals, newly constructed housing units are added to each annual sample as they are added to the housing unit frame.

These new construction cases are sampled using the same TE and are assigned the same rotation and month as the initial sample. This means that, in any given month, a small number of new construction cases are on their first NCVS interview even though their sample rotation indicates they should be on their second through sixth interview. If only the seventh interview remains when a new construction case is added to the sample, the case is not sent out for interviews.

GQ frame

Each GQ within the GQ frame is assigned to a measure of size based on the expected number of units or persons within the GQ. This measure of size transforms the GQ size into the number of housing unit equivalents within the GQ. As with the housing unit frame, a systematic sample is selected within each PSU after the GQ measures are sorted by state, county, census tract, percentage of GQs within the block that are college dormitories, and census block. Each sample hit identifies a census block, and the hit string is four housing unit equivalents.

The GQ frame TE is the same as the housing unit frame TE from the same sampling year, but divided by three to select three years of sample rather than one. Each hit string is then assigned to one of the six rotations that will begin interviewing during the following three years.

Around six months before a GQ is scheduled to begin interviewing, interviewers conduct GQ listings. This involves interviewing a contact person at each sampled GQ to confirm the GQ type and create a listing of units within the GQs. After this listing is created, a subsample of units is selected within each sampled GQ. This subsample is a systematic sample using a subsampling TE that is designed to select the desired number of units based on the expected number of units within the GQ. If the actual number of units within the GQ is much higher than expected, further subsampling is done to ensure the NCVS interviewer workload is not too high within the GQ.

Sample sizes

The NCVS first- and second-stage sample sizes are designed to be large enough to produce annual national violent crime estimates with a coefficient of variation (CV) of no more than 5 percent. As of 2016, the NCVS sample is also designed to produce state-level estimates with CVs of no more than 10 percent using up to three years of data within the largest 22 states.

The expected CV is a function of expected crime rate and expected standard deviation. The expected crime rate is based on previous NCVS results, and the expected standard deviation is a function of the expected crime rate and the design effect. The design effect is a measure of the difference in theoretical variance estimates between a complex sample design and a simple random sample. In other words, the design effect quantifies the effect on the variance due to the NCVS sample and panel design. The NSR PSUs have within-PSU and between-PSU variance, and the SR PSUs only have within-PSU variance. Therefore, the overall design effect is different for SR and NSR PSUs.

Based on the number of SR and NSR PSUs and the expected design effect within each PSU type, there is a target number of completed person interviews within each state that will result in the desired CVs for national and state-level violent crime estimates. These target interviews are allocated to the sample PSUs within the state proportionate to the relative size of the population the PSU represents. For SR

19

PSUs, it is based on the population size of the PSU because the PSU only represents itself. For NSR PSUs it is based on the population size of the NSR sampling stratum, which includes the sampled PSU and all of the non-sampled PSUs within the stratum.

The second-stage sample size is based on these target completed interviews, then adjusted to account for expected occupied housing unit rates, housing unit response rates, average persons age 12 and over within occupied housing units, and within-household response rates. These adjustments result in an expected sample size of around 240,000 housing units per year during a normal survey year. These sample sizes may increase or decrease as response rates and persons per household change.

Each annual sample uses the same sampling TE for both the housing unit and GQ frame, and this sampling TE is determined by the number of units in the combined housing unit and GQ frames at the time of sampling. This results in an overall NCVS sample that has roughly the same distribution of housing units and GQs as the units in the housing unit and GQ frames. Around 2 percent of the NCVS sample comes from the GQ frame.

Special Sample Situations

During a normal year of NCVS data collection, the NCVS sample is equally divided into seven groups, or rotations, of cases on their first through seventh NCVS interview. These different rotations have historically seen different response rates and crime reporting rates due to a variety of factors. All rotations are combined together to calculate NCVS estimates, and year-to-year estimate changes are valid as long as the sample distributions are consistent. This section describes some situations where the sample distributions are not consistent and users should use caution when making year-to-year comparisons.

Phase-in and Phase-out of new and old sample designs

Every ten years, the NCVS sample is redesigned to select PSUs based on population data from the most recent decennial census. As part of these redesigns, sampling methods and frames may change to use more current data, use more sophisticated methods, or address changes in the scope of the NCVS. One notable difference between the 2000 and 2010 design NCVS samples is the sampling frames. The 2000 design NCVS sample was based on four different sampling frames—Unit, GQ, Area, and Permit—and the frames were assigned by census block. The Unit, GQ, and Permit frames provided all addresses within blocks with mostly city-style addresses where building permit offices monitored new construction. The Area frame covered the remainder of the sample area and required interviewers to conduct regular address listing operations to determine the units within the blocks. The 2000 design NCVS sample was designed only for national estimates, and all within-PSU sample selection was completed once at the beginning of the decade.

Another notable difference between the 2000 and 2010 designs is the sampling weights. The 2000 design NCVS was a self-weighting sample, which means that all sample cases had the same sampling base weight. The 2010 design NCVS is not self-weighting, which allows for oversampling in the top 22 states and in areas with low expected response rates. This variability in base weights may increase the design effects, but larger sample sizes counteract this increase so the national CVs are generally unchanged.

The 2015 NCVS marks the first year of the overlap between the 2000 design and the 2010 design NCVS samples. The first-stage sample remained the same, but new rotations within the sample

counties were selected from the 2010 design frame for the 2000 design counties that are also in sample in the 2010 design (continuing counties). The 2010 sample is larger due to the new state-level estimate goal, so this resulted in a sample distribution that had slightly more sample on the first two interviews than in interviews three through seven. During a normal year, the NCVS sample has approximately one seventh of the sample on interview one through seven. Due to the phase-in of new 2010 PSUs, this distribution was not even during the 2016 and 2017 NCVS.

In 2016, the NCVS began interviewing in 2010 design PSUs and producing NCVS estimates for geographies defined by the 2010 census. Within continuing counties, the existing 2000 design cases continued interviewing through December 2017, when the last 2000 design rotation was phased out. In outgoing counties, all NCVS interviewing stopped after December 2015. In new counties, all NCVS interviewing began in January 2016. The 2018 NCVS is the first year of fully phased-in 2010 design sample.

Sample Reductions and Reinstatements

Sample distributions may change during steady state years due to sample reductions and reinstatements. Sample reductions occur when NCVS sample cases that have already been interviewed previously are not sent out for later interviews. This usually occurs when budget constraints limit the amount of sample that can be fielded during a fiscal year. If later budgets allow, this reduced sample may be reinstated to finish the remaining NCVS interview cycles. Whenever this occurs, special weighting adjustments are usually applied to account for differences between the actual and expected interview number for the reinstated sample cases. If sample reductions or reinstatements occur, more details will be provided in the annual Source and Accuracy Statement.

Chapter 3. Data collection

Field operations

Interviewer training

The Census Bureau generally employs between 1,000 and 1,500 NCVS interviewers, or field representatives (FRs), across the United States. Training for NCVS interviewers consists of classroom and on-the-job training. Initial training for interviewers 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 protecting respondent confidentiality, gaining respondent cooperation, answering respondent questions, proper survey administration, use of systems to collect and transmit survey data, NCVS concepts and definitions, and completing simulated practice NCVS interviews.

The NCVS procedures and concepts taught in initial training are also regularly reinforced for experienced NCVS interviewers. This information is received via monthly written communications, ongoing feedback from observations of interviews by supervisors, and monthly performance and data quality feedback reports.

In the past, experienced interviewers received different types of refresher training. In 2011, a refresher training experiment was conducted. Two evaluations of this experiment are available on the Census Bureau's website:

https://www.census.gov/srd/papers/pdf/rrs2013-07.pdf

https://www.census.gov/srd/papers/pdf/rrs2014-02.pdf

Interview mode

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).

The initial interview with a sample household is typically 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. The NCVS interviews conducted in subsequent enumeration periods are generally conducted by telephone. In 2016, approximately 49 percent of all interviews were conducted by telephone.

The expected interview time is approximately 25 minutes. Interview times for respondents with no victimization incidents to report can be expected to be shorter than this. In addition, 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.

Detailed statistics showing the relationship between interview mode, and respondent and household characteristics can be found in the BJS publication, *Interviewing Conditions in the National Crime Victimization Survey, 1993-2013* which is accessible at the following location: https://www.bjs.gov/content/pub/pdf/icncvs9313.pdf.

Dangerous addresses

Interviewers play an important role in gathering NCVS information from sample persons or addresses. However, the safety of the interviewers 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 by an authorized official with knowledge from the field that the status of the address has changed.

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, interviewers 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.

Consent and confidentiality

All of the data for NCVS are collected by BJS under the authority of Title 34 U.S. Code (U.S.C.) § 10132. In addition, BJS is required to keep all personally identifying information about respondents strictly confidential, under the authority of Title 34 U.S.C. §§ 10134 and 10231.

Likewise, all information collected as part of the NCVS is held in strictest confidence under Title 13, U.S.C. 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:

"The U.S. Census Bureau is required by law to protect your information. The Census Bureau is not permitted to publicly release your responses in a way that could identify you. We are conducting this survey for the Bureau of Justice Statistics of the United States Department of Justice under the authority of law (Title 13, United States Code, Section 8). The Bureau of Justice Statistics is authorized to collect this survey information by law (Title 34, United States Code, Section 10132). Federal law protects your privacy and keeps your answers confidential (Title 13, United States Code, Section 9 and Title 34, United States Code, Sections 10134 and 10231). Per the Federal Cybersecurity Enhancement Act of 2015, your data are protected from cybersecurity risks through screening of the systems that transmit your data."

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

Reference period

Respondents are asked to report crime experiences occurring in the 6 months preceding the month of interview. Research has shown that respondents generally 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 (Bhandari & Wagner, 2006).

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. 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. Over-reporting 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.

Persons included

Three types of respondents are included in the NCVS:

- household respondents
- individual respondents
- proxy respondents

Certain NCVS questions are designed for different respondents. Questions relating to crimes affecting the household are asked only once during each enumeration period. The respondent for these questions is referred to as the household respondent. The household respondent must be at least 18 years old, and is the person most knowledgeable about the household.

Other questions in the NCVS are considered self-response questions and relate to crimes affecting persons and not the household. In other words, each household 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 household member. This person is referred to as a proxy respondent. The use of proxy respondents is discussed in more detail in the next section.

Use of proxy interviews

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. However, under certain circumstances proxy interviews are permitted. 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 age 12 or older 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 must 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 parent or guardian 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.

Because of the limited circumstances under which they are allowed, proxy interviews typically account for less than 5 percent of NCVS interviews. Table 2 presents a breakout of 2016 NCVS proxy interviews by reason.

Table 2. 2016 NCVS proxy interviews by reason

Reason	Number	Percent
Age 12-13	1,485	12%
Physically/mentally unable	4,926	39%
Unavailable	6,171	49%
Other	10	< 1%

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2016.

In 2016, 5.61% of all interviews were conducted by proxy. Of these, about 49 percent were because the respondent was temporarily absent and would not be back to the address until after the interview closeout date. In addition, about 39 percent were because the respondent was physically or mentally unable, and about 12 percent were because a parent or guardian refused to allow the interviewer to interview their child age 12 or 13.

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 various 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 generally account for about 4 percent of households.

Spanish and alternative language support

Currently, the NCVS questionnaire is available in both English and Spanish. (NCVS introductory letters are also available in English, Spanish, Chinese, Korean, and Vietnamese.) For Spanish and other languages, interviewers 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 interviewer if he/she is fluent in the respondent's language.

If finding a suitable interpreter is difficult while making contact, field supervisors may help interviewers obtain assistance. If a suitable interpreter cannot be located, proxy respondents are not allowed for the NCVS.

Typically, non-English language interviews account for about 4 percent of NCVS interviews, with Spanish typically accounting for about 3.6 percent, Chinese accounting for about 0.2 percent, and all other languages accounting for 0.1 percent or less of all NCVS interviews.

Reinterview

Reinterview is a major feature of both the quality assurance (QA) and the missed crimes estimation program. The NCVS QA reinterview uses two approaches: random and supplemental (supervisor discretion) to validate interviewer performance. The missed crimes estimation program uses the data from the QA program to estimate household and person level missed crimes.

The random reinterview approach consists of selecting a sample of each interviewer's work to review over the data collection cycle. The supplemental approach allows supervisors to identify additional interviewers or cases for review throughout the cycle. Reinterview requires that a supervisor or experienced interviewer re-contact respondents at a sample of previously-interviewed households. Reinterviewers verify that the original interviewer contacted the correct sample unit, determined the correct household composition, and classified noninterview households correctly.

Reinterviewers also verify the household roster and tenure, ensure specific questions are covered, and re-ask a subset of the crime screener questions. Normally, the original household respondent will answer the household crime screener questions during reinterview. However, a proxy household member age 18 or older is allowed if the household respondent is not available. 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.

Some reasons for maintaining the NCVS reinterview program include:

- detect and deter data falsification
- provide information on interviewer performance
- provide information on household and person missed crimes
- call attention to a need for revising procedures or instructions
- determine whether refresher training is required on some particular phase of the survey

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, either at one of the Census Bureau's phone centers or at a regional office. 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.

Data quality feedback

Another component of the data quality program is monthly feedback. In 2011, the Census Bureau implemented a series of field performance and data quality indicators. Previously, high response rates were the primary measure of interviewer performance. The data quality indicators are tracked through the Census Bureau's expanded Performance and Data Analysis (Giant PANDA) tool, and monthly reports provided to the field. Under the revised performance structure, interviewers are monitored on the following:

- response rates (household, person, and the current supplement in the field)
- time stamps (the time it takes to administer the screener questions on the NCVS-1 or the crime incident questions on the NCVS-2)
- overnight starts (interviews conducted very late at night or very early in the morning)
- late starts (cases not started until the 15th or later in the interview month)
- absence of contact history records (cases missing records of contact attempts with the household and/or persons within the household)
- quality of crime incidents (changes made to the location, presence, or theft data items on the NCVS-2 during post-processing coding operations)

Noncompliance with these indicators results in supervisor notification and follow-up with the interviewer. 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 interviewer training, or removal of the interviewer from the survey.

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.

The survey instruments associated with each of these stages—the NCVS-500 Control Card, the NCVS-1 Basic Screen Questionnaire, and the NCVS-2 Crime Incident Report—may be found on the BJS webpage dedicated to the NCVS: https://www.bjs.gov/index.cfm?ty=dcdetail&iid=245.

NCVS-500 Control Card

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

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.

In addition to the crime screener questions, the NCVS-1 also contains socio-demographic questions including employment (asked of individual household members) and household income (asked of the household respondent). Beginning in July 2016, additional socio-demographic questions were added to the NCVS-1 including veteran status, citizenship, gender identity and sexual orientation, and disability questions were moved from the NCVS-2 Crime Incident Report to the NCVS-1.

NCVS-2 Crime Incident Report

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. The exception is if the respondent reports a series crime. This is when a specific type of crime has occurred more than five times in the past six months and the respondent cannot remember the details of each individual incident. In this case, the interviewer documents the entire series of incidents with a single CIR.

Within the CIR, questions are asked in topical sections: location and presence, attack/threat/injury/medical care, emotional toll, actions against the offender, offender

characteristics, attempted and completed thefts, property damage, reporting to police, activity at the time of incident and time/money lost, series of crimes, hate crime, and a written summary of the incident.

Supplements

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. See Chapter 1 for descriptions of the most recent NCVS supplements.

Nonresponse

As in most surveys, three types of missing data are in the NCVS: household nonresponse, person nonresponse, and item nonresponse.

Household 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.

At the closeout of an interview period, each sample household is assigned a final interview outcome code, indicating whether the household was an interview or noninterview. An NCVS sample address is 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.

Annual rates for all three types of noninterviews are included in the NCVS Source and Accuracy Statement, which is available as part of the Codebook beginning in 2016.

Appendix D lists all valid interview outcome codes, including sample household noninterview types.

Person nonresponse

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, weighting, and imputation. This is discussed in further detail in Chapters 4 and 5.

An eligible household member is 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 only to a household member, not to the entire sample household.

Item nonresponse

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 key items that are important in modeling crime. Annual rates of item nonresponse can be found in the NCVS Source and Accuracy Statement, which is available as part of the Codebook beginning in 2016.

Chapter 4. Post-data collection processing

The NCVS data undergo several processing and editing steps after being pulled from the interview instrument. The editing and coding operation is done first on a monthly basis, before any other processing. The remainder of the monthly processing is then performed. This includes pre-editing, matching to the unit control file (UCF), industry and occupation coding, and reformatting. When three 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, family structure recoding, weighting, recoding, and stripping type of crime (TOC) code 90 from the public use file. All processing stages are explained in further detail later in this section.

Instrument data reformat

The first step in processing is to extract data from the interview instrument and reformat the data for use in subsequent data processing steps. These data are also used to produce status reports to aid in the monitoring of data collection progress and data quality.

Coding operations

After the data are pulled from the interview instrument and reformatted for processing, several coding operations take place.

Manual editing and coding

Editing and coding is a manual operation in which clerks at the Census Bureau's National Processing Center (NPC) review every case containing at least one crime incident report; at least one crime reported through items 44a and 45a on the NCVS-1, which are the crime catchall questions; or any respondent race and ethnicity entry of "Other-specify." Identified issues 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, catchall crime response, or respondent race and ethnicity entry of "Other-specify." 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. The initial reviewer can correct or accept the data or add a referral code to send the case to headquarters staff for further review. 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 crime catchall entries as well as race and ethnicity entries of "Other-specify" are also reviewed and, when possible, recoded to one of the pre-coded categories.

When all of the cases have gone through the initial review, incidents then go through a second review (i.e., 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 not 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 1,300 cases are on the file sent to the NPC, of which about 500 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. They also review all recodes for race and ethnicity entries of "Other-specify," and classifications for crime catchall questions. After headquarters staff have completed their review, the editing and coding file for the month is closed out.

Industry and occupation coding

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. This process uses the 4-digit North American Industry Classification System (NAICS) industry and 4-digit Standard Occupational Classification (SOC) occupation codes to assign the appropriate I&O code to the respondent's place of work.

Geographic identification coding

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.

Middle school teacher recode

Some of the codes used to classify occupations are collapsed, and include multiple occupations. For example, during the I&O coding process, coders classify elementary school teachers and middle school teachers with the same occupation code. However, the 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.

Pre-edit

During the pre-edit stage of monthly processing, basic data integrity checks are done to flag inconsistencies that may be evident in the data that need to be addressed at a later stage in the edit process. 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 Type A noninterview 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.

UCF match

When the pre-edit is complete, each NCVS record is matched to the 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 [Federal Statistical Research Data Centers](#) (FSRDCs).

Adjust/reformat

The adjust/reformat process is completed monthly after performing the pre-edit and the match to the UCF. The adjust/reformat process applies corrective action to the edit failures identified during the pre-edit process. During this process, programmatic corrections are made to the data based on errors flagged during the NCVS computer pre-edit. This adjustment process sometimes involves classifying the household as a Type A noninterview.

Range check

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.

Consistency edit and imputation

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.

The Census Bureau's traditional, sequential hot-deck procedure is then used to impute any remaining missing or rejected values for selected characteristics of interviewed and some noninterview persons. Individual characteristics imputed for interviewed and Type Z noninterview persons include age, sex, Hispanic origin and race. Income, tenure, land use, farm sales and type of living quarters are also imputed for household interviews, Type A noninterviews, and Type B noninterviews. For each missing value, the procedure assigns a value reported for a person with similar characteristics, also known as a donor record, or based on other known information about the individual. For some items subject to imputation, sex and age specifically, both the original and allocated values are kept on the data file. In addition, variables on the incident report may be allocated based on consistency edits with other variables. Annual imputation rates can be found in the NCVS Source and Accuracy Statement, which is available as part of the [Codebook](#) beginning in 2016.

Blanking edit

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."

Type of crime classification

Each crime incident is assigned a type of crime (TOC) code that depends on the entries in the incident report. Each crime 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 considered the 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.

Period-to-period recount

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—six 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.

Industry and occupation code edits

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.

Family structure recode

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. This recode creates a "family structure" variable that is added to the NCVS data file.

Weighting

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.

TOC 90 removal

All TOC 90s, unwanted sexual contact crimes, are also stripped from the public use file. Examples of unwanted sexual contact crimes include grabbing and fondling, with or without force. Incidents of this type are not classified into any crime category and are not considered crimes for the purposes of the NCVS.

Public use file creation

After all the previous processing steps are completed, a public use file is created. Blanking, collapsing, and topcoding are applied to the public use file to meet the data confidentiality requirements defined by the Census Bureau Disclosure Review Board (DRB).

Quality measures and internal files

The final steps in processing include the creation of quality measures and indicators of nonsampling error within the NCVS data, and the creation of final internal NCVS data files for use in the FSRDCs.

Chapter 5. Weighting

Household, person, and victimization data from the NCVS sample are adjusted to give annual and bi-annual estimates of crime experienced by the U.S. population age 12 or older. Household and person weights are first adjusted to account for any subsampling that occurs within large GQs. The nonresponse weighting adjustment then allocates the sampling weights of nonresponding households and persons to respondents with similar characteristics. The ratio adjustment reduces 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.

This section discusses the various components of NCVS weights, which are summarized in Table 3.

Table 3. Components of NCVS weights

Components of the NCVS weights	Household-level estimates			Person-level estimates		
	Household	Victimization	Incident	Person	Victimization	Incident
Base weight	×	×	×	×	×	×
GQ subsampling adjustment	×	×	×	×	×	×
Household nonresponse	×	×	×	×	×	×
Within-household nonresponse				×	×	×
Ratio adjustment	×	×	×	×	×	×
Bounding adjustment		×	×		×	×
TIS adjustment		×	×		×	×
Series crime adjustment		x	x		x	x
Multiple victim adjustment						×

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2016.

Base weights

The original NCVS base weight for each housing unit or GQ is the inverse of the probability of selection for that case. The overall probability of selection is the product of the first- and second-stage probabilities of selection. Base weights differ across PSUs due to the different probabilities of selection of the NSR PSUs and the different sampling rates within the PSUs.

All cases within a PSU have the same first-stage probability of selection. The second-stage sampling is done annually, and all cases within a PSU have the same second-stage probability of selection within each annual sample. This results in all cases within a PSU having the same base weight within each annual sample. However, second-stage sampling rates vary slightly across annual samples, so some base weight variability is expected within PSUs when multiple annual samples are combined to produce NCVS estimates.

Each annual sample is divided so half begins interviewing in the first rotation of the year and the other half begins interviewing in the second rotation of the year. All base weights are adjusted such that each half of the annual sample will have weighted sums equal to one seventh of the population. This ensures that the combination of all seven interview rotations in a given six month period will sum to the population estimate.

Weighting adjustments

If all eligible units in the sample responded to the survey and reported crimes only within the reference period, the sampling base weights would produce unbiased estimates with reasonably low variance. However, nonresponse and nonsampling errors are expected in all sample surveys, and the following post-data-collection weighting adjustments minimize their impact on the NCVS estimates. All of these adjustments are completed on six months of response data at a time.

GQ subsampling

Some units in the GQ frame are subsampled because the observed GQ size is much larger than expected. During the estimation procedure, units within these GQs must receive a GQ subsampling adjustment (also known as the weighting control factor) to account for the change in the probability of selection. For example, a GQ address expected to have 25 units but found at the time of GQ listing to contain 75 units, could be subsampled at the rate of 1 in 3 to reduce the interviewer's workload. Each of the units in this case would be given a weighting control factor of 3. To limit the effect of this adjustment on the variance of sample estimates, these GQ subsampling factors are limited to a maximum value of 4. This introduces a small amount of bias in order to avoid large increases in variance.

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. 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. 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 occupied households are tabulated separately for each nonresponse adjustment cell. The weight for this purpose is the base weight multiplied by any GQ subsampling factor. The household nonresponse factor (HHNRF) is computed as—

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

where A is the weighted count of interviewed households and B is the weighted count of noninterviewed households.

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

- Noninterview Cluster (NICL) code, which groups PSUs together within the top 22 states or census division by SR/NSR status, expected response rate, and expected crime reporting rates
- 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)
- urbanicity (only if CBSA/MSA status is not the principal city within a CBSA/MSA)
- race of the household reference person (white only or all other races)
- interview number groups for the address (1-2, 3-4, or 5-7)

These variables are cross-classified to make original cells for the household nonresponse adjustment. Cells are collapsed when the nonresponse factor is greater than or equal to 2.0 or if the cell contains fewer than 30 responding households. 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. The interview number groups are collapsed first, then reference person race, then urbanicity, and finally CBSA/MSA status. There is no collapsing across NICL.

Within-household nonresponse
A household is considered a response if at least one person within the household completes the NCVS interview. The interviewer then attempts to interview all persons age 12 and older within the household, but some persons within the household may be unavailable or refuse to participate in the survey. This within-household nonresponse adjustment allocates the weights of nonresponding persons to respondents. The starting weight for all persons within responding households is the same household-level base weight multiplied by any GQ subsampling factor and the household nonresponse adjustment factor.

The within-household nonresponse factor (WHNRF) is computed as—

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

where C is the weighted count of interviewed persons and D is the weighted count of noninterviewed persons within interviewed households.

If nonrespondents' crime victimizations are significantly different from respondents' crime victimizations, there could be nonresponse bias in the NCVS estimates. To reduce nonresponse bias, the within-household nonresponse adjustment cells are formed by characteristics that are

correlated with response and crime victimization rates. This includes: top 22 states/region, age, sex, race/ethnicity, and relationship to household reference person (self/spouse or all others). These variables are cross-classified in different ways, depending on household relationship, to create 54 cells within each state or region.

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. Age is collapsed first, then sex, and finally race/ethnicity. There is no collapsing across household relationship or state/region.

Ratio Adjustment

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 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.

The NCVS ratio adjustment has three high-level steps: (1) person coverage, (2) person iterative raking, and (3) household coverage. The person coverage step controls person weights to American Community Survey (ACS) estimates by census region, CBSA/MSA status, and household reference person race (black-only versus all others). This is a ratio adjustment where the numerator is the ACS estimate within the adjustment cell, and the NCVS population estimate is the denominator. If any initial factors are less than 0.5 or greater than 2, the reference person races are collapsed together to create one cell for the region and CBSA/MSA status.

The person iterative raking step controls person weights to independent population controls[3] by different demographic groups and geography levels. The first dimension of raking creates cells within state cells for the top 22 states and within census region for the remaining states. The Northeast and Midwest regions are combined to increase sample sizes. Within each geography, the first dimension cells are created by sex, age (12-29, 30-49, and 50+), and race/ethnicity (Hispanic, non-Hispanic black-only, and non-Hispanic all other races). The second dimension of the person iterative raking step divides the entire sample by race/ethnicity (Hispanic, non-Hispanic white-only, non-Hispanic black-only, and non-Hispanic other races), sex, and age (12-13, 14-15, 16-17, 18-19, 20-21, 22-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, and 75+).

If any initial cells in the person iterative raking step have fewer than 30 respondents or adjustment factors less than 0.5 or greater than 2, they are collapsed with adjacent cells until these criteria are met. The first dimension collapses first across race, then across age, and finally across sex. The

---

[3] The independent population controls are prepared by projecting forward the population figures derived from Census 2010, using information from a variety of other sources that account for births, deaths, and net migration. Prepared in this manner, the controls are themselves estimates but are derived independently of the NCVS and provide useful information for adjusting sample estimates.

second dimension collapses first across age, then sex, and finally race/ethnicity. Due to the lower geography, the first dimension is expected to have a high number of cells collapse. The second dimension is expected to have no cells collapse.

The person iterative raking step performs a ratio adjustment on the first dimension cells then the second dimension cells, and repeats this process up to 250 times until the weighted sums are within 0.1 percent of the population control in all cells in both dimensions. After this is complete, the sums of person weights will match the independent controls for all collapsed dimension cells.

The household coverage step is the final step of the ratio adjustment and applies only to household weights. First, the ratio adjustment factor for the household principal person[4] is applied to the household weight. Then the household weights are controlled to independent estimates of housing units. The independent housing unit estimates include vacant and occupied housing units, but the NCVS produces estimates of only occupied housing units. Therefore, the sampling base weight for vacant units are included in NCVS weighted sums when calculating the household coverage adjustment factor.

Bounding

Telescoping occurs when respondents report events that fall outside of the period of interest (Neter & Waksberg, 1964). Telescoping causes over-reporting and often happens in surveys when respondents are asked to recall all events within a given period. The NCVS asks respondents to recall all incidents that occurred during the previous 6 months. Prior to 2006, the first NCVS interview was a bounding interview and was not used in estimates, to avoid potential telescoping bias. In 2006, the first of the seven NCVS interviews in new sample areas was used in estimates, in conjunction with a bounding adjustment for the first interview, to avoid telescoping bias. All of the first NCVS interviews have been included in the estimates with a bounding adjustment since 2007.

The bounding adjustment factor is defined as—

$$BFAC = \frac{\sum_{t=2}^{7} CR_t}{6 * CR_1}$$

where CR_t is the unbounded weighted crime rate in the previous twelve months for cases on TIS t. Only sample cases where TIS = 1 on the rotation chart receive this bounding adjustment. All other sample cases do not receive a bounding adjustment (BFAC=1).

The bounding adjustment is calculated for property crimes and violent crimes separately and applied to the final household or person weights to create the victimization weight for each incident.

Time-in-sample

The NCVS has a known time-in-sample (TIS) effect, wherein respondents in earlier interviews report more incidents than respondents in later interviews. See Biderman & Cantor (1984) for more information on TIS bias. The NCVS sample usually consists of a roughly equal number of cases on

---

[4] The household principal person is the wife, if there is one, and is the reference person otherwise.

each TIS group, so this TIS bias does not impact the estimates. However, adding a large number of sample cases in one particular TIS may impact comparisons to incident rates of prior survey years.

The TIS adjustment factor is designed to prevent bias in the estimates when large groups of sample cases begin (or resume) interviewing in a TIS that is not what they are assigned in the rotation chart. This only occurs in the special situations of sample reductions and reinstatements and the phase-in of new design sample. For example, in the first half of 2016 all sample within new 2010 design counties was on their first interview even though the rotation had the sample evenly divided among the seven TIS groups. The bounding adjustment accounted for respondent telescoping only for sample cases in rotation chart TIS 1. The TIS adjustment accounted for respondent telescoping for all other new sample cases in that time period.

The TIS adjustment factor is defined as—

$$TISFAC = \frac{CR_T}{CR_t}$$

where CR_T and CR_t are the unbounded crime rates from the previous four quarters for sample cases within TIS groups *T* (from the rotation chart) and *t* (the actual TIS). If *T=t* for the sample case, there is no TIS adjustment (TISFAC = 1). During years when the sample is evenly distributed across TIS groups, there is no TIS adjustment factor for any NCVS sample cases.

As with the bounding adjustment, the TIS adjustment factor is calculated for property crimes and violent crimes separately, then applied to the appropriate victimization weight based on the type of crime for each incident.

Series crimes

When a respondent reports a series crime, the interviewer completes one incident report for all incidents with the details of the most recent incident. In order to count all instances of this incident, the victimization weight is multiplied by the number of times (up to 10) the incident occurred.

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.

Multiple victims

If every victimization had one victim, the incident weight would be the same as the victimization weight. Because incidents sometimes have more than one victim, the incident weight is the series victimization weight divided by the number of victims in the incident.

This concept does not apply to property crimes, so the incident weight is always the same as the series victimization weight for property crimes. For violent crimes, the number of victims in an incident can be greater than the number of persons in the household. It is also possible for one violent crime incident report in the NCVS to be a series crime and have multiple victims.

Replicate weights

The NCVS uses 160 replicate weights to produce variance estimates that reflect the complex sample design and weighting adjustments. To produce these replicate weights, the sampling base weights are

multiplied by 160 different replicate factors (which are described below) to produce replicate base weights. Each set of replicate base weights is subjected to the same weighting adjustments described in the previous section to produce 160 sets of final replicate weights for households, persons, series victimizations, and incidents. By applying the weighting adjustments to each replicate, the final replicate weights reflect the impact of the weighting adjustments on the variance.

Replicate factors are created differently based on the PSU type, and the following subsections describe the methods for each.

NSR replicate factors

Replicate factors based on the Balanced Repeated Replication (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.

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 for the top 22 states and within census division for the remaining states. When possible, pseudo strata contain two NSR PSUs. However, some areas contained an odd number of NSR PSUs that resulted in one pseudo stratum containing three NSR PSUs. One state contained only one NSR PSU, so that pseudo stratum could also only contain one NSR PSU.

For pseudo strata containing one or two PSUs, the PSUs (or randomly assigned halves of the one PSU) within each pseudo stratum were randomly assigned to one of two panels of the replicate. The entire pseudo stratum was also assigned to one row from a Hadamard matrix of order 160. The same replicate factors were then assigned to all cases within each PSU (or half) using the formulas—

$$f_1(r,c) = 1 - P_1 H_{c,r} + H_{c,r}$$

$$f_2(r,c) = 1 - P_1 H_{c,r}$$

where $f_p(r,c)$ is replicate factor r for panel p (1 or 2) using Hadamard matrix row c, $H_{c,r}$ is the value of the Hadamard matrix at row c and column r, and P_1 is the proportion of the pseudo stratum population assigned to panel 1 (see Wolter, 1985, for more information). For the pseudo stratum with just one PSU, P_1=0.5 and the replicate factors are 1.5 or 0.5.

For pseudo strata containing three PSUs, the PSUs were randomly assigned to one of three panels of the replicate. The entire pseudo stratum was assigned to two subsequent rows (denoted *c* and *c+1*) from a Hadamard matrix of order 160. The same replicate factors were then assigned to all cases within each PSU using the formulas –

$$f_1(r,c) = 1 + \frac{1}{4}\left(3H_{c,r}\hat{P}_3 + H_{c+1,r}\sqrt{3}\left(2\hat{P}_2 + \hat{P}_3\right)\right)$$

$$f_2(r,c) = 1 + \frac{1}{4}\left(3H_{c,r}\hat{P}_3 + H_{c+1,r}\sqrt{3}\left(\hat{P}_3 - 2\hat{P}_1\right)\right)$$

$$f_3(r,c) = 1 + \frac{1}{4}\left(3H_{c,r}\left(-\hat{P}_1 - \hat{P}_2\right) + H_{c+1,r}\sqrt{3}\left(\hat{P}_2 - \hat{P}_1\right)\right)$$

SR replicate factors

Because the variation of the SR PSUs comes entirely from selecting units within the PSU, the successive difference method is used to assign replicate factors within SR PSUs. 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 method used for the NCVS.

To assign replicate factors, the SR sample is sorted by PSU and within each PSU by the same order that was used to select the original systematic sample. All the units within each sample hit are assigned two rows of a Hadamard matrix of order 160, always skipping rows 1 and 80 because they contain all 1s. This assignment begins with rows (2,3) assigned to the first hit, rows (3,4) assigned to the second hit, and so on until rows (159,160) and (160,2). The assignment is repeated, increasing the difference between rows by one, so the next hit is assigned rows (2,4), then rows (4,6), and so on until rows (157,159) and (159,2). Once all possible combinations are assigned moving forward, the entire process is repeated in reverse. This is all repeated until the entire sample is assigned two rows.

Using the assigned rows (*c* and *d*) of the Hadamard matrix, each sample hit is assigned replicate factors as—

$$f_r = 1 + 2^{-\frac{3}{2}}H_{c,r} - 2^{-\frac{3}{2}}H_{d,r}$$

where f_r is replicate factor *r* using Hadamard matrix rows *c* and *d*. This formula yields replicate factors of approximately 1.7, 1.0, or 0.3.

Final replicate weights

The replicate factors are just the first step in creating final replicate weights. The second step is creating 160 replicate base weights by multiplying the sampling base weight for each sample unit by the 160 replicate factors calculated above. Each set of replicate base weights is then adjusted using the same weighting adjustments described above. All replicate weighting adjustments use the same cell collapsing as the primary weights rather than repeating cell collapsing for each replicate.

This method accounts for the effect on the variance of the other weighting factors. Recalculating the noninterview and 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.

Supplements

Design weights specific to the NCVS supplements are produced so population and incident totals and rates may be estimated for the supplements. Supplements are at the person level and are not intended to produce household estimates. If every person that responded to the NCVS would also respond to the NCVS supplement, the supplement weight would be the same as the final person weight for NCVS. However, persons may respond to the NCVS but not the supplement, so supplement weighting adjustments ensure that supplement weights continue to represent the entire population of interest.

All supplement base weights are set to the final person weight from the NCVS. Then supplement weighting adjustments may include supplement-specific nonresponse adjustments and/or additional ratio adjustments. Each supplement uses different weighting adjustments, and the supplement Source and Accuracy Statement will describe each in detail. There are no separate victimization weights for supplements because the longer reference period (one year) and longer time between supplement interviews make bounding unnecessary.

Supplement replicate weights are calculated by multiplying the final supplement weights by the NCVS replicate factors. There are no additional steps to apply supplement-specific weighting adjustments to the supplement replicate weights.

Chapter 6. Estimation

The NCVS data allow users to produce estimates of crimes and crime rates. This section describes methods used for generating these point estimates and variance estimates.

Total crime estimates

Point estimates of crime victimizations and crime incidents ($\hat{t}$) are calculated using the formula –

$$\hat{t} = \sum_{i \in S_I} w_i$$

where S_I is all incidents reported by sample units within the domain and time period of interest, and w_i is the appropriate weight for sample case *i*. For crime victimizations, use the series victimization weight. For crime incidents, use the incident weight. If six months of data are used for the calculation, the estimate will represent the crime victimizations or incidents within that six month period. For annual estimates, use the full year of data.

Crime rate estimates

The BJS calculates NCVS crime rate estimates as the number of victimizations per one thousand people. This estimate is not a proportion because it is possible for one household or person to have multiple victimizations within any time period. The formula for crime rate estimates is –

$$\hat{r} = 1000 * \frac{\hat{t}}{\hat{N}}$$

where $\hat{t}$ is the victimization estimate defined above, and $\hat{N}$ is the population estimate. The population estimate is calculated as –

$$\hat{N} = \frac{\sum_{i \in S} w_i}{M/6}$$

where *S* is all sample households or persons within the domain and time period of interest, w_i is the household weight (for property crimes) or the person weight (for violent crimes), and *M* is the number of months of NCVS data included in the estimates. All NCVS weights are designed so that six months of data produce full population estimates, so adjustments are needed whenever fewer or more than six months of data are used to produce an estimate.

Variance estimates

The NCVS estimates come from a sample, so they may differ from figures from an enumeration of the entire population using the same questionnaires, instructions, and enumerators. For a given estimator, the average squared difference between estimates based on repeated samples and the estimate that would result if the sample were to include the entire population is known as sampling error. The sampling error quantifies the amount of uncertainty and bias in an estimate as a result of selecting a sample.

The sample estimate and its standard error enable one to construct a confidence interval, which is a range about a given estimate that has a specified probability of containing the average result of all possible samples. A 90-percent confidence interval ranges from 1.645 standard errors below the sample estimate to 1.645 standard errors above the estimate.

This means that if all possible samples were surveyed under essentially the same general conditions and using the same sample design, then approximately 90 percent of the individual sample confidence intervals would include the average estimate of all possible samples.

Standard errors may also be used to perform hypothesis testing, a procedure for distinguishing between population parameters using sample estimates. The most common type of hypothesis is that the population parameters are different. An example of this would be comparing the victimization rates of men to the victimization rates of women.

Direct variance estimates

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.

The weighting section described the creation of replicate weights that are used to calculate variance estimates. The Census Bureau calculates the direct variance estimate of the total crime estimate $\hat{t}$ using Fay's Balanced Repeated Replication (BRR) method (Fay, 1989) as –

$$var(\hat{t}) = \frac{1}{R(1-k)^2} \sum_{r=1}^{R} (\hat{t}_r - \hat{t})^2$$

where $\hat{t}$ is the point estimate described above, $\hat{t}_r$ is the point estimate using replicate r weights, R is the number of replicates (160 for NCVS), and k is the Fay coefficient (0.5 for NCVS).

The variance estimate for the crime rate is similarly calculated as –

$$var(\hat{r}) = \frac{1}{R(1-k)^2} \sum_{r=1}^{R} \left(\frac{\hat{t}_r}{\widehat{N_r}} - \frac{\hat{t}}{\widehat{N}} \right)^2$$

where $\widehat{N_r}$ is the population estimate using replicate r weights.

Generalized variance functions

Every year, the Census Bureau produces parameters for generalized variance functions (GVFs) that estimate the variance of any crime count estimate based on the value of the estimate. To do this, estimates and their relative variance are fit to a regression model using an iterative weighted least

squares procedure where the weight is the inverse of the square of the predicted relative variance. This is done using all crime estimates listed in Table 1 (except serious violent crimes) for the overall population and all domains in Table 4.

Table 4: NCVS Domains for GVF calculations

Violent Crime Domains	Property Crime Domains
Urban	Urban
Suburban	Suburban
Rural	Rural
Household Income < $25,000	Household Income < $25,000
$25,000 ≤ Household income < $50,000	$25,000 ≤ Household income < $50,000
$50,000 ≤ Household income < $100,000	$50,000 ≤ Household income < $100,000
Household income ≥ $100,000	Household income ≥ $100,000
Reported to police	Reported to police
Not reported to police	Not reported to police
Age 12-24	Reference person age 12-34
Age 25-49	Reference person age 35-49
Age 50+	Reference person age 50+
Race black-only	Reference person race black-only
Race not black-only	Reference person race not black-only
Hispanic	Reference person Hispanic
Non-Hispanic	Reference person non-Hispanic
Married/widowed	Single-person household
Single/divorced/separated	Multiple-person household
Male	Tenure = owner
Female	Tenure = renter/no cash rent

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2017.

The resulting parameters of this model allow data users to calculate the variance of any estimate with a GVF of the formula—

$$\widehat{var}(\hat{t}) = a\hat{t}^2 + b\hat{t} + c\hat{t}^{3/2}$$

where $\hat{t}$ is the point estimate of victimizations, $\widehat{var}(\hat{t})$ is the estimated variance of the point estimate, and *a*, *b*, and *c* are parameters estimated through the regression model. See Treat (2017) for more information on how GVFs are computed.

There are four different sets of parameters for the four different types of crime estimates:

1. Overall personal crimes
2. Domain personal crimes
3. Overall property crimes
4. Domain property crimes

Parameter set #1 is used for the overall person crime estimates. These are the person crime estimates by crime category for the whole population, not disaggregated by any victim or offender characteristics, or any variable related to reporting to police. For example, parameter set #1 should be used to generate variances for total aggravated assault rates.

Parameter set #2 is used for the person crime domain estimates. These are the person crime estimates disaggregated by victim or offender characteristics, or any variable related to reporting to police. For example, parameter set #2 should be used to generate variances for aggravated assault rates among women.

Parameter set #3 is used for the property crime estimates for the whole population. These are the property crime estimates by crime category for the whole population, not disaggregated by any household characteristics or any variable related to reporting to police. For example, parameter set #3 should be used to generate variances for total motor vehicle theft rates.

Parameter set #4 is used for the property crime domain estimates. These are the property crime estimates disaggregated by household characteristics or any variable related to reporting to police. For example, parameter set #4 should be used to generate variances for motor vehicle theft rates in urban areas.

When computing variances for "all crimes" (person and property together), parameter set #3 should be used. When the person and property estimates are combined and disaggregated by victim, household, or offender characteristics, or any variable related to reporting to police, parameter set #4 should be used. Refer to Appendix F for illustrations of how to use the GVF parameters to estimate variances.

Year-to-year correlations

When making comparisons between two different estimates, it is important for the variance of the difference to account for any correlation between the two estimates. The variance for the difference between two estimates ($\hat{x}$ and $\hat{y}$) is –

$$var(\hat{x} - \hat{y}) = var(\hat{x}) + var(\hat{y}) - 2\rho\sqrt{var(\hat{x})var(\hat{y})}$$

where ρ is the correlation coefficient between the two estimates. When comparing estimates between two domains that do not overlap, the correlation coefficient is zero. Due to the panel design of the NCVS, there is some correlation between different years of the same estimates.

The Census Bureau calculates year-to-year correlation coefficients for all overall crime estimates for the crime classifications listed in Table 1 (except serious violent crimes[5]). The correlations change each year. Correlations between the current year and the previous two years can be found in the NCVS Source and Accuracy Statement, which is available as part of the Codebook beginning in 2016. Year-to-year correlations for 2014 and 2015 are included in Appendix G. Correlations for earlier years are available in the 2013 NCVS Technical Documentation.

---

[5] The BJS uses aggravated assault correlations when calculating year-to-year change estimates of serious violence.

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

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

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

The estimator for the year-to-year correlation for a general statistic θ between time t and s is defined as—

$$\hat{\rho}(\hat{\theta}_t, \hat{\theta}_{t-s}) = \frac{\widehat{cov}(\hat{\theta}_t, \hat{\theta}_s)}{\sqrt{\widehat{var}(\hat{\theta}_t)}\sqrt{\widehat{var}(\hat{\theta}_s)}}$$

which is defined in terms of the sample replicate covariances and variances that are functions of the replicate estimator θ_t and θ_s at times time t and s. That is—

$$\widehat{cov}(\hat{\theta}_t, \hat{\theta}_{t-s}) = \frac{1}{R(1-k)^2}\sum_{r=1}^{R}(\hat{\theta}_{t,r} - \hat{\theta}_t)(\hat{\theta}_{s,r} - \hat{\theta}_s)$$

where R is the number of replicates (160 for NCVS) and k is the Fay coefficient (0.5 for NCVS). The NCVS year-to-year correlations are calculated using estimates of the crime rate and not estimates of crime totals.

Appendix A. Terms and definitions

This section provides the definitions of several terms used within this technical document. It includes terms related to statistics, sample design, crime, and victimization.

2010 sample design. The sample design implemented to select the samples used to calculate estimates from 2016 through 2025. The sample design is referred to as 2010 because it followed directly after the 2010 Census and used many of the data products from the 2010 Census to select an up-to-date sample.

Bias. The formal definition of bias of an estimator $\hat{\theta}$ of some statistic θ 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 the 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 primary sampling units 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), and the BRR formulas used by the NCVS are defined by Fay (1989).

Bounding. A process to ensure that previously reported incidents are not reported again in the enumeration that follows. Incidents reported in the prior interview are used to confirm duplicate incidents in the current interview. Bounding provides a more accurate measure of criminal victimization within NCVS sample households.

Central city/Balance/Urban/Rural (CBUR). A geographic identifier that indicates whether a sample block is in a central city of a MSA, the balance of an urbanized area, an urban cluster, or a rural cluster (i.e., outside the urbanized area or cluster).

Coefficient of variation (CV). The coefficient of variation is a standardized measure of dispersion of a probability distribution or frequency distribution. It is often expressed as a percentage, and is defined as the ratio of the standard deviation to the estimate. That is, $\sqrt{var(\hat{\theta})} / \hat{\theta}$. It is also the square root of the relative variance. The BJS considers an estimate to be unreliable (and should be interpreted with caution) if the CV exceeds 50 percent.

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.

Core-Based Statistical Area (CBSA)/Metropolitan Statistical Area (MSA) Status. The classification used by the Census Bureau to delineate whether a metro area is (1) located within a principal city or cities of the MSA, (2) located within a MSA and *not* in a principal city or cities of the MSA, or (3) not located within a MSA.

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 percent coverage, then the frame and the sample design include 75 percent 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, serious violent crimes, and property crimes. See "Incident" and "Victimization."

> **Personal crime.** Rape, sexual assault, 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, robbery, and assault. Threats of harm include attempted attacks and attempted sexual assaults by means of verbal threats.

> **Serious violent crime.** Rape, sexual assault, robbery, and aggravated assault are considered serious violent crimes.

> **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/index.cfm?ty=tdtp&tid=9>

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."

Design effect. A measure of the difference in theoretical variance estimates between a complex sample design and a simple random sample. Calculated as the complex sample design variance divided by the simple random sample variance estimate with the same sample size.

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).

Field representative (FR). The Census Bureau's term for an interviewer.

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

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."*

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.

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. There are three basic types of nonresponse in the NCVS: household nonresponse, which is when no interviews are obtained from anyone in an eligible household; person nonresponse, which is when an interviewer fails to obtain an interview from at least one, but not all, of the eligible persons in a household; and item nonresponse, which is when a respondent completes part of the questionnaire but does not answer one or more individual questions.

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. The NCVS selects one PSU per first-stage stratum, and at least two sample units are needed to calculate a variance. Therefore, the first-stage strata are grouped, with two (and sometimes three) PSUs within each pseudo strata.

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

Reference person. For NCVS, the reference person is a responsible adult household member who is not likely to leave the household on a permanent basis. 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, 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. That is, $v(\hat{\theta}) / \hat{\theta}^2$. It is also equal to the square of the coefficient of variation.

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 less than 1.0. A SR PSU represents itself and no other PSUs because it is the only PSU in its stratum. A NSR PSU represents itself and the other PSUs 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. The 2000 design NCVS 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).

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).

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 systematic sample design. The main ideas of replication are outlined by Fay and Train (1995).

Systematic Sample. A type of probability sampling method in which sample members are selected from a sorted list according to a random starting point (start-with) and a fixed periodic interval (take-every).

Telescoping. Telescoping occurs when respondents report crime events that fall outside of the reference period.

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.

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. 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 is the expected value of the squared difference between an estimator θ and its expected value. That is, $v(\hat{\theta}) = E\left(\hat{\theta} - E(\theta)\right)^2$.

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. Chronology of NCVS question changes since January 2000

Date	Description of change
July 2000	Hate crime questions revised; Disability questions added in outgoing rotations only
January 2001	Disability questions added for all persons reporting criminal incidents
July 2001	Added cyber-crime questions
January 2003	Race and ethnicity of respondents questions revised based on OMB directive; Educational attainment questions revised; Included hate crime question on the public use file
January 2004	New questions added to related to special entry/gated community and the number of handguns/firearms stolen; Revised disability questions
July 2004	Removed cyber-crime questions and added household-level identity theft questions
July 2005	Pregnancy question for female respondents added
January 2007	Disability questions revised for compatibility with the American Community Survey (ACS); Victim-offender relationship categories revised
January 2008	Disability questions revised for compatibility with the ACS
July 2008	Vandalism questions deleted
July 2008	Emotional impact questions added
January 2012	Identity theft questions deleted; replaced with ITS
January 2012	Race and ethnicity of offender questions revised
July 2016	New questions on citizenship, veteran status, sexual orientation, and gender identity added; Disability questions now asked of all respondents; Household income answer categories expanded

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2000–2016.

Appendix C. NCVS Supplements Conducted, 2000-2017

Supplement name	Dates conducted	Universe of interest	Topic
Identity Theft Supplement (ITS)	2016, 2014, 2012, 2008	NCVS respondents age 16 and older	Identity theft
Police-Public Contact Survey (PPCS)	2015, 2011, 2008, 2005, 2002	NCVS respondents age 16 and older	Prevalence of contact with police
School Crime Supplement (SCS)	2017, 2015, 2013, 2011, 2009, 2007, 2005, 2003, 2001	NCVS respondents ages 12–18 enrolled in school	School-related victimization
Supplemental Fraud Survey (SFS)	2017	NCVS respondents age 18 and older	Financial fraud
Supplemental Victimization Survey (SVS)	2016, 2006	NCVS respondents age 18 and older in 2006; Expanded to age 16 and older beginning with 2016 cycle	Stalking, harassment or unwanted contact/behavior
Workplace Risk Supplement (WRS)	2002	NCVS respondents age 16 or older currently employed or employed at least 2 weeks in the last 6 months	Workplace violence

Source: Bureau of Justice Statistics, National Crime Victimization Survey, 2000–2017.

Appendix D. Interview Outcome Codes

Outcome Code	Description
201	Completed interview (no Type Zs)
203	Sufficient partial - no more follow-up needed
204	Sufficient partial - follow-up needed
Type A	
213	Type A - Language problems
216	Type A - No one home
217	Type A - Temporarily absent
218	Type A - Refused
219	Type A - Other occupied
Type B	
225	Type B - Temporarily occupied by persons with usual residence elsewhere
226	Type B - Vacant - regular
227	Type B - Vacant - storage of furniture
228	Type B - Unfit or to be demolished
229	Type B - Under construction, not ready
230	Type B - Converted to temporary business or storage
231	Type B - Unoccupied site for mobile home, trailer, or tent
232	Type B - Permit granted, construction not started
233	Type B - Other
Type C	
240	Type C - Demolished
241	Type C - House or trailer moved
242	Type C - Outside segment
243	Type C - Converted to permanent a business or storage
244	Type C - Merged
245	Type C – Condemned
247	Type C - Unused line of listing sheet
248	Type C - Other
258	Type C - Unlocatable sample address
259	Type C - Unit does not exist or is out of scope

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

Appendix E. SAS code for direct variance estimation

The NCVS public use files are located on the National Archive of Criminal Justice Data (NACJD) website <http://www.icpsr.umich.edu/icpsrweb/NACJD/NCVS/>. The SAS program below demonstrates how to use these datasets to produce a direct variance estimate of population, crimes, and crime rates using the replicate weights. To improve readability, the variable names in this program do not match the variable names on the dataset. Data users should refer to the codebook for actual variable names.

```
**********************************************

*** Calculating estimates of 2016 collection year motor vehicle theft in Urban
*** areas (principal city within MSA).
***
*** Input datasets:
***     house - household record-type file.
***             1 record per HU with all data and weights.
***     incident - incident-level extract file.
***                1 record per incident with all data and weights.
***
*** Note:  Replicate HH weights are not adjusted for the annual file, so the
***        variance formula in this program applies the annual adjustment to only
***        the replicate sums.
**********************************************;

* Sum the full sample (annual adjusted) and replicate (not adjusted) HH weights;
proc means data=house sum noprint;
   where msa_status = '1';    * subset to urban households;
   var hhwgt_1yr hhrepwgt1-hhrepwgt160;
   output out=house_sums sum=pop pop1-pop160;
run;


* Sum series weight and replicate series weights;
proc means data=incident sum noprint;
   * subset to motor vehicle thefts committed within the U.S in urban areas;
   where msa_status = '1' and toc_new in ('40','41') and outside_us ne '1';
   var series_weight seriesrepwgt1-seriesrepwgt160;
   output out=incident_sums sum=crimes crimes1-crimes160;
run;


data final_estimates;
   merge house_sums incident_sums;
   array reppop{160} pop1-pop160;
   array repcrime{160} crimes1-crimes160;
   crime_rate = crimes/pop;      * create crime rate estimate as ratio of sums;
   do i=1 to 160;  * loop through replicates to sum squared differences;
      var_pop + (4/160)*((reppop{i}/2)-pop)**2;
      var_crimes + (4/160)*(repcrime{i}-crimes)**2;
      var_rate + (4/160)*((2*repcrime{i}/reppop{i})-crime_rate)**2;
   end;
   crime_rate = crime_rate*1000; * adjust crime rate to crimes per 1000 persons;
   var_rate = var_rate*(1000**2); * same adjustment to variance of crime rate;
run;


proc print;
   var pop var_pop crimes var_crimes crime_rate var_rate;
   format pop var_pop crimes var_crimes comma20. crime_rate var_rate 5.1;
run;
```

Appendix F. Illustrations of GVF variance estimation

The illustrations in this appendix use the GVF parameters from the 2014 and 2015 NCVS, but the methods are the same for every survey year. For current GVF parameters, refer to the Source and Accuracy Statement for the current survey year.

<u>Illustration 1</u> – Crime victimization total

If the estimated number of motor vehicle thefts (a property crime) for all households (an overall estimate) is 400,000, then the overall property crime GVF parameters estimate the standard error of that estimate as—

$$s_t = \sqrt{a\hat{t}^2 + b\hat{t} + c\hat{t}^{\frac{3}{2}}}$$
$$= \sqrt{-0.00020489(400{,}000^2) + 3{,}510(400{,}000) + 2.048(400{,}000^{3/2})}$$
$$= 43{,}466$$

where $\hat{t}$ is the point estimate for total crimes and a, b, and c are GVF parameters for overall property crimes. The 95-percent confidence interval is $400{,}000 \pm 1.96(43{,}466)$.

Illustration 1 – Crime victimization total	
Number of motor vehicle thefts	400,000
a parameter	-0.00020489
b parameter	3,510
c parameter	2.048
Standard error	43,466
95-percent confidence interval	314,806 to 485,194

<u>Illustration 2</u> – Crime victimization rate

If there are approximately 42,000,000 households in urban areas (a domain subset) and 1.1 million burglaries occurred in urban areas, then the burglary rate in urban areas is:

$$\hat{r} = \frac{\hat{t}}{\hat{N}} = \frac{1{,}100{,}000}{42{,}000{,}000} = 0.026$$

and the standard error is—

$$s_r = \sqrt{\frac{b\hat{r}(1-\hat{r})}{\hat{N}} + \frac{c\hat{r}(\sqrt{\hat{r}} - \hat{r})}{\sqrt{\hat{N}}}}$$
$$= \sqrt{\frac{3{,}105(0.026)(1 - 0.026)}{42{,}000{,}000} + \frac{2.282(0.026)(\sqrt{0.026} - 0.026)}{\sqrt{42{,}000{,}000}}}$$
$$= 0.0018$$

where $\hat{t}$ is the point estimate for total crimes, $\hat{N}$ is the number of households in the domain, and b and c are the GVF parameters for domain property crimes.

If the estimated crime rate is represented as the number of crimes per 10,000 persons (by multiplying by 10,000), then use this formula to calculate the standard error of the rate (before multiplying by 10,000) and multiply the resulting standard error by 10,000. A similar approach may be used to produce estimates of percentages by multiplying by 100. Illustration 3 demonstrates how to calculate variances for estimates of crimes per 1,000 persons.

Illustration 2 – Crime victimization rate	
Number of burglaries in urban areas	1,100,000
Number of households in urban areas	42,000,000
Burglary rate in urban areas	0.026
a parameter	-0.0002222
b parameter	3,105
c parameter	2.282
Standard error	0.0018
95-percent confidence interval	0.023 to 0.030

Illustration 3 – Difference between two domain estimates

The standard error of the difference between two sample estimates is approximately equal to—

$$s_{x_1-x_2} = \sqrt{s_{x_1}^2 + s_{x_2}^2 - 2\hat{\rho}s_{x_1}s_{x_2}}$$

where s_{x_1} and s_{x_2} are the standard errors of the estimates x_1 and x_2 and $\hat{\rho}$ is the correlation coefficient between them. Note that x_1 and x_2 could be totals, rates, or percentages.

When comparing estimates between two different domains in the same year, there is no sample overlap, and the correlation coefficient is 0. Estimates that are more than three years apart are assumed to be uncorrelated. Estimates less than three years apart have a correlation coefficient that is generated by the Census Bureau and provided in the annual Source and Accuracy Statement.

If 130 million males experienced a rate of 2.2 robberies per 1,000 persons in 2015, and 140 million females experienced a rate of 2.1 robberies per 1,000 persons in 2015, then the domain person crime GVF parameters estimate the standard error of each crime rate as—

$$s_m = 1,000 * \sqrt{\frac{3,682(0.0022)(1-0.0022)}{130,000,000} + \frac{13.251(0.0022)(\sqrt{0.0022}-0.0022)}{\sqrt{130,000,000}}} = 0.420$$

$$s_f = 1,000 * \sqrt{\frac{3,682(0.0021)(1-0.0021)}{140,000,000} + \frac{13.251(0.0021)(\sqrt{0.0021}-0.0021)}{\sqrt{140,000,000}}} = 0.397$$

and the standard error of the difference between them is then:

$$s_{m-f} = \sqrt{0.420^2 + 0.397^2 - 0} = 0.578$$

where s_m, s_f, and s_{m-f} are the standard errors of the male estimate, female estimate, and their difference respectively. The formula for each domain standard error is adjusted to produce standard errors of an estimate of crimes per 1,000 persons, but the formula for the difference between the estimates requires no such adjustment.

Illustration 3 – Difference in Crime Rates Between Two Domains			
	Males	Females	Difference
Domain population count	130,000,000	140,000,000	-
Robbery rate (per 1,000 persons)	2.2	2.1	0.1
b parameter	3,682	3,682	-
c parameter	13.251	13.215	-
Standard error	0.420	0.397	0.578
95-percent confidence interval	1.4 to 3.0	1.3 to 2.9	-1.0 to 1.2

Illustration 4 – Difference between two years

To calculate the standard error of an estimate's change between two consecutive years, use each year's GVF parameters to calculate the standard error for the previous year estimate (s_{prev}) and the current year estimate (s_{curr}). For example, assume the 2014 overall violent crime count was 5.4 million, and the 2014 GVF parameters estimated the standard error $s_{prev} = 327{,}958$. Also, assume the 2015 overall violent crime count was 5.0 million, and the 2015 GVF parameters estimated the standard error $s_{curr} = 311{,}964$. Then the estimated change is 400,000 fewer crimes with standard error of the change calculated as—

$$s_{prev,curr} = \sqrt{327{,}958^2 + 311{,}964^2 - 2(0.101)(327{,}958)(311{,}964)} = 429{,}198.$$

Illustration 4 – Difference in Crime Totals Between Two Years			
	2014	2015	Difference
Total Violent Crimes	5,400,000	5,000,000	-400,000
a parameter	-0.00050049	-0.000482	-
b parameter	4,230	4,840	-
c parameter	7.914	7.618	-
Correlation coefficient	-	-	0.101
Standard error	327,958	311,964	429,198
95-percent confidence interval	4,757,202 to 6,042,798	4,388,551 to 5,611,449	-1,241,228 to 441,228

Appendix G. NCVS year-to-year correlations, 2014 and 2015

Crime Category	2013-2014	2012-2014	2014-2015	2013-2015
Total Personal Crimes	0.001	0.217	0.109	-0.104
Crimes of Violence	-0.005	0.204	0.101	-0.097
Completed	-0.079	0.017	0.089	-0.032
Attempted	0.029	0.077	0.002	-0.066
Rape/Sexual Assault	0.074	-0.134	0.020	-0.102
Rape	0.092	-0.022	0.016	0.020
Completed	-0.072	-0.048	0.029	0.087
Attempted	0.021	0.048	-0.018	-0.055
Sexual Assault	0.134	-0.155	0.021	-0.200
Robbery	-0.043	-0.047	-0.067	-0.029
Completed	-0.051	-0.028	-0.219	-0.031
With Injury	-0.037	-0.102	-0.061	0.034
Without Injury	-0.005	0.017	-0.236	-0.033
Attempted	0.015	0.018	0.143	-0.040
With Injury	-0.109	0.021	0.129	0.089
Without Injury	-0.005	-0.002	0.079	0.076
Assault	-0.024	0.184	0.147	-0.003
Aggravated	0.060	-0.010	-0.094	0.096
Completed with Injury	0.109	-0.046	-0.077	0.087
Attempted	0.015	0.003	0.125	0.058
Simple	0.011	0.109	0.072	-0.021
Completed with Injury	-0.121	0.043	0.269	0.018
Attempted	-0.018	0.067	-0.003	-0.009
Purse Snatching/Pocket Picking	-0.048	0.035	-0.057	-0.171
Completed	0.029	0.008	-0.119	-0.085
Attempted	-0.033	0.034	0*	0*
Total Property crimes	0.323	0.277	0.403	-0.037
Burglary	0.094	-0.057	0.150	-0.040
Completed	-0.005	-0.022	0.135	-0.108
Forcible Entry	-0.027	0.102	0.075	-0.024
Unlawful Entry without Force	0.013	-0.067	0.125	-0.098
Attempted Forcible Entry	0.065	0.088	0.016	-0.025
Motor Vehicle Theft	0.160	0.072	0.152	0.058
Completed	0.174	0.044	0.163	0.086
Attempted	-0.052	0.239	0.078	0.041
Theft	0.243	0.261	0.436	-0.016
Completed	0.264	0.268	0.434	-0.035
Less than $50	0.126	0.029	0.341	-0.119
$50 - $249	0.111	0.176	0.210	0.099
$250 or more	0.148	-0.056	0.291	0.021
Amount not available	0.166	0.077	0.060	0.074
Attempted	-0.033	0.012	-0.070	-0.010

*No attempted purse snatching or pocket pickings were reported in 2015. Therefore, there is no correlation when comparing attempted purse snatchings or pocket pickings between 2015 and another year. To compute difference estimates between the zero attempted purse snatchings/pocket pickings for 2015 and the number from a different year, a correlation of zero should be used in the standard error formula.

References

American Association for Public Opinion Research. (2011). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 7th edition*. AAPOR.

Bhandari A., & Wagner, T. Self-reported utilization of health care services: improving measurement and accuracy. *Medical Care Research and Review*, 63 (2006), pp. 217-235

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.

Cochran, W.G. (1977). *Sampling Techniques*. New York: John Wiley & Sons.

Fay, Robert E. (1989). Theory and Application of Replicate Weighting for Variance Calculations. In *Proceedings of the Survey Research Methods Section, American Statistical Association*, 212-217.

Fay, R. E., & Train, G. (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.

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.

Ohlsson, Esbjorn (2000). Coordination of PPS Samples Over Time. In *The Second International Conference on Establishment Surveys, American Statistical Association*, 255-264.

Treat, J. (2017). *Generalized Variance Function Methodologies for the National Crime Victimization Survey, v2.0*. U.S. Census Bureau memorandum for the record, signed March 2, 2017.

Wolter, K. M. (1984). An investigation of some estimators of variance for systematic sampling. *Journal of the American Statistical Association 79,* 781–790.

Wolter, K. M. (1985). *Introduction to Variance Estimation*, New York: Springer-Verlag.