USER'S GUIDE TO NATIONAL CRIME VICTIMIZATION SURVEY (NCVS) DIRECT VARIANCE ESTIMATION

Bonnie Shook-Sa, MAS

G. Lance Couzens, BS

Marcus Berzofsky, DrPH

RTI International
3040 East Cornwallis Road
Research Triangle Park, NC 27709

CONTENTS

<u>Appendices</u>

LIST OF TABLES

INTRODUCTION

The National Crime Victimization Survey (NCVS), sponsored by the Bureau of Justice Statistics (BJS), estimates the frequency and describes characteristics of criminal victimization in the United States. When calculating NCVS estimates, researchers must take into account the complex stratified, four-stage sample design and analysis weights. Stratification, clustering, and variation in analysis weights all affect the variances of survey parameters, and inappropriately accounting for these factors during estimation can lead to invalid results.[1]

Two broad methods exist for calculating variances of estimates from complex sample designs. One method is generalized variance functions (GVFs), which model the design-consistent variances for multiple survey estimates to obtain GVF parameters. Using the formulas and parameters from these models, GVFs allow users to calculate approximations of variances without knowledge of the sample design. This method has traditionally been used for NCVS estimation. The GVF equations and parameters that have historically been used to calculate NCVS standard errors can be found in *Appendix B*. Furthermore, BJS has made available to NCVS analysts an online tool, the NCVS Victimization Analysis Tool (NVAT), that computes estimates and standard errors based on GVFs for several common estimate types and domains of interest.[2]

The second method is to estimate variances directly using software that accounts for complex sample designs. Two direct variance techniques are Taylor series linearization (TSL) and balanced repeated replication (BRR). Compared to BRR, TSL has been shown to be a more straightforward method for NCVS estimation. The design variables required to calculate TSL estimates are already available on NCVS public use files (PUFs), whereas the BRR weights must be computed for single and pooled years of NCVS data.[3]

The purpose of this guide is to provide documentation for calculating direct variance estimates from the NCVS using TSL. The guide includes the three key NCVS estimate types

---

[1]   Cochran, W. G. (1977). *Sampling techniques*. New York, NY: Wiley.
[2]   Available at http://www.bjs.gov/index.cfm?ty=nvat
[3]   Williams, R., Heller, D., Couzens, G. L, Shook-Sa, B., Berzofsky, M., Smiley-McDonald, H., & Krebs, C. (2014). *Evaluation of direct variance estimation, estimate reliability, and confidence intervals for the National Crime Victimization Survey*. Prepared for the Bureau of Justice Statistics, Washington, DC.

found in most BJS reports: victimization totals, proportions, and rates. *Section 1* summarizes the data structure of the NCVS PUFs. *Section 2* outlines how to prepare NCVS public use data files for direct variance estimation and describes newly available NCVS analytic files that can be used for direct variance estimation. *Section 3* discusses direct variance estimation for each estimate type. *Section 4* recommends means for identifying low-quality estimates.

Appendix A contains example SAS, SPSS, and Stata code for calculating the series adjustment. *Appendix B* provides the GVF equations and parameters that were historically used to calculate NCVS standard errors. *Appendices C–F* provide detailed examples of direct variance estimation in SAS, SUDAAN, SPSS (Complex Samples Package), and Stata, respectively.

User's Guide Terms and Notation

The following two tables define frequently used terms and outline the notation in this NCVS user's guide.

Terms Used in This Guide	
Generalized variance functions (GVFs)	GVFs use formulas and parameters from models of direct variance estimates that allow users to calculate approximations of variances without knowledge of the sample design. This is the current method of variance estimation for the National Crime Victimization Survey (NCVS).
Taylor series linearization (TSL)	TSL is the method of direct variance estimation documented in this user's guide. TSL provides variance estimates of linear and nonlinear statistics through Taylor series expansions of linear functions.
Balanced repeated replication (BRR)	BRR is a method of direct variance estimation in which a set of replicate weights is developed on the basis of the complex sample design. Replicate weights are used to provide design-consistent variance estimates of statistics.
Personal crimes	The NCVS defines personal crimes as those committed against an individual. The NCVS classifies personal crimes into five crime types: rape/sexual assault, robbery, aggravated assault, simple assault, and personal theft. Classifications are based on the *V4529* variable on the NCVS public use file (PUF).
Property crimes	The NCVS defines property crimes as those committed against households. The NCVS classifies property crimes into three crime types: household burglary, motor vehicle theft, and theft. Classifications are based on the *V4529* variable on the NCVS PUF.
Primary sampling units (PSUs)	In a household survey with a cluster sample design,* PSUs are typically geographic areas that are the first level of selection in the sample. On the NCVS PUF, the actual PSUs are not available because of disclosure concerns. Instead, PSUs are approximated with the pseudo-stratum† (*V2117*) and half-sample§ (*V2118*) codes.
Series victimizations	Series victimizations occur when an NCVS respondent recalls at least six criminal incidents of a similar nature but cannot recall the dates and other details of the individual incidents well enough to report them separately. Each series victimization is represented as a single record on the incident-level file, and the number of incidents linked to that series is recorded. Weight adjustments are needed to appropriately account for series victimizations when computing NCVS estimates. The series count (*serieswgt*) and series-adjusted victimization weight (*newwgt*) are included on the modified analysis files.

Terms Used in This Guide	
Victimization summary file	The victimization summary file, which contains data regarding reported victimizations needed for the calculation of victimization rates, is calculated from the incident-level file. For personal crimes, the victimization summary file contains one record for each person and reporting period. For property crimes, the victimization summary file contains one record for each household and reporting period. For both types of crime, the victimization summary file contains the number of victimizations with the specified characteristics and the victimization weight. This file is merged with the person- or household-level file for the calculation of victimization rates. Victimization summary files need to be computed only for the calculation of victimization rates for incident-level characteristics not included on the modified person- and household-level files.
Victimization adjustment factor	The victimization adjustment factor is used in the calculation of victimization rates. It is defined as the ratio of the victimization weight to the household weight (for property crimes) or the person weight (for personal crimes). The victimization adjustment factor is included on the modified person- and household-level files, so this adjustment factor does not need to be calculated by the analyst.

*Cluster sampling methods group sampling units into subgroups called clusters. Rather than selecting sampling units directly, analysts select clusters for inclusion in the sample. Cluster sampling is used when a list of sampling units is not available at the time the sample is selected and when the analyst needs to control the geographic distribution of sampled units for logistical data collection purposes (Cochran, W. G. [1977]. *Sampling techniques*. New York, NY: Wiley.).

†Pseudo-strata are available on the NCVS PUFs and approximate the true sampling strata.

§Half-sample codes are available on the NCVS PUFs and approximate the true PSUs.

Notations Used in This Guide	
i	Represents National Crime Victimization Survey (NCVS) households. These can be identified on the household-level file on the basis of the household identification number (*IDHH*).
j	Represents NCVS individual respondents within households *i*. These can be identified on the person-level file on the basis of the person identification number (*IDPER*).
k	Represents reporting periods (i.e., *YEARQ*) for households *i* and individual respondents *j*.
l	Represents victimization records for respondent *j* in household *i* and reporting period *k*. Each record on the NCVS incident-level file is associated with a victimization record *l*.
D	Represents one or more domain characteristics of interest in the calculation of NCVS estimates. For victimization totals and proportions, domains can be defined on the basis of crime types (e.g., violent crimes, property crimes), characteristics of victims (e.g., age, sex, household income), or characteristics of the victimizations (e.g., victimizations reported to police, victimizations committed with a weapon present). Domains could also be a combination of all of these types of characteristics. For the calculation of victimization rates, domains are defined on the basis of characteristics of the victims.
A_a	Represents level *a* of covariate *A*. Covariate *A* is defined in the calculation of victimization proportions and represents the characteristic for which the analyst wants to obtain the distribution of victimizations in domain *D*.
C	Crime type *C* is defined in the calculation of NCVS victimization rates. It is the personal or property crime for which the analyst wants to obtain a victimization rate.

SECTION 1. DATA STRUCTURE

NCVS PUFs are available through the National Archive of Criminal Justice Data within the Interuniversity Consortium for Political and Social Research (ICPSR) at the University of Michigan. NCVS data, which are available in SAS, SPSS, Stata, R, and ASCII formats, are organized in four files: an address-level file,[4] a household-level file, a person-level file, and an incident-level file.[5] Because the NCVS employs a 6-month retroactive reference period for reporting crimes, data are available in two file structures: collection-year files and data-year files. Collection-year files report crime incidents on the basis of when the interviews were conducted (as opposed to when the incidents actually occurred). Data-year files report crime incidents on the basis of when the incidents occurred, regardless of when the interviews were conducted. Different sets of survey weights are provided for each type of file. Under the collection-year approach, only 12 months of interviews are needed for annual estimation. With the data-year approach, annual estimates cannot be made until 18 months of interviews have been conducted, making collection-year data more timely. For most outcomes, the difference between the collection-year and data-year estimates is not statistically or substantively different. Therefore, collection-year data are more commonly used in BJS analyses and reports. Although this user's guide will introduce both collection- and data-year weights, the examples will focus on collection-year estimation.

1.1   Household-Level File

The household-level file contains one record for each sampled household in the NCVS per reporting period. It contains data from the household screening interview, which assesses whether a household experienced any property crimes during the previous 6 months. The household-level file also contains characteristics of (1) the household's surrounding area, such as the census region and Metropolitan Statistical Area status, and (2) the principal and reference persons within the household.

---

[4]   The address-level file is not used in the calculation of NCVS estimates and is therefore not further documented in this section.

[5]   ICPSR includes both an incident file and an incident extract file. The incident extract file includes the associated household- and person-level variables, eliminating the need to merge these variables from the person- and household-level files. This document assumes that the analyst is using the incident extract file.

Because households stay in the NCVS sample for a total of 3 years—reporting seven times at 6-month intervals—most households appear on the annual household file more than once. For this reason, both the household identification number (*IDHH*) as well as the year and quarter indicator (*YEARQ*) must be used to uniquely identify households by reporting period when merging the household file with other NCVS data files. Household-level estimates for the collection year use the collection-year household weight (*WGTHHCY*), whereas estimates for the data year are based on the data-year household weight (*WGTHHDY*). The household file is most commonly used in the calculation of property victimization rates (see **Section 3.1.3** for more details).

1.2   Person-Level File

The person-level file contains data for each household member aged 12 or older in responding NCVS households. Each record corresponds to a sampled person within a reporting period. Data come from the personal screening interviews, which are administered to all eligible and participating household members. The screening interview determines whether a person experienced a personal victimization during the previous 6 months.

Like the household file, the annual person-level file contains multiple records for most respondents. In addition to the year and quarter indicator (*YEARQ*), the person identification number (*IDPER*) must be used to uniquely identify persons within reporting periods when merging with other NCVS data files. The household identification number (*IDHH*) is available on the person-level file to allow merges with corresponding household-level data. Person-level estimates for the collection year use the collection-year person weight (*WGTPERCY*), whereas estimates for the data year are based on the data-year person weight (*WGTPERDY*). The person file is most commonly used in the calculation of personal victimization rates (see **Section 3.1.3** for more details).

1.3   Incident-Level File

In most cases, the incident-level file contains one record for each victimization reported by NCVS respondents. It contains both property crimes reported by the household respondent (i.e., household burglary, motor vehicle theft, and theft) and personal crimes reported by any NCVS respondent (i.e., rape/sexual assault, robbery, aggravated assault, simple assault, and

personal theft). In addition, the incident-level file contains the corresponding household- and person-level variables from the household and person files. The incident file contains data to classify victimizations on the basis of crime type as well as details of each victimization drawn from the incident report (e.g., persons present, victim-offender relationship, weapon use). If the respondent reports six or more criminal incidents of a similar nature but cannot recall specific details of each incident, the incidents are collapsed into a single record on the incident-level file and the total victimization count is recorded. These types of victimizations, called series victimizations, are further discussed in *Section 2*.

Because respondents can report multiple incidents in a single interview, the incident file contains multiple records for some respondents and no records for respondents who did not report any criminal victimizations. The incident file can be linked to the household file with the household identification number (*IDHH*) and the year and quarter indicator (*YEARQ*). For personal crimes, the incident file can be linked to the person file with the person identification number (*IDPER*) and the year and quarter indicator (*YEARQ*). Estimates based on the incident file for the collection year are based on the collection-year victimization weight (*WGTVICCY*), whereas estimates for the data year are based on the data-year victimization weight (*WGTVICDY*).[6] Victimization weights are used to calculate victimization totals, proportions, and property and personal victimization rates (see *Section 3.1* for more details).

---

[6] The victimization weights available on the NCVS PUF do not account for series victimizations. A modified series-adjusted victimization weight (*newwgt*) can be derived and used in analyses that include series victimizations, as discussed in *Section 2.1*.

SECTION 2. PREPARING DATA FILES FOR DIRECT VARIANCE ESTIMATION

Victimization totals and proportions are estimated on the basis of data reported by victimized households or persons, not the total population of households or persons. This allows estimation from a single file (the incident file) and analysis weight (the victimization weight). Because the incident file contains data only for persons and households reporting victimizations, several steps are needed to prepare the incident file for direct variance estimation to ensure that variance estimates appropriately reflect the NCVS design. *Section 2.1* outlines the preparation of the incident file for direct variance estimation of victimization totals and proportions.

Calculating victimization rates requires knowledge about the total population and the victimized population. Victimization rates are calculated by taking the ratio of the number of victimizations to the total population and multiplying this ratio by 1,000. The numerator of the victimization rate is estimated from the incident-level file, using the victimization weight. For property crimes, the denominator is calculated from the household-level file with the household weight. For personal crimes, the denominator is calculated from the person-level file with the person weight. Because estimates of victimization rates are based on two files and two sets of weights, which current software packages cannot accommodate, pre-processing is needed before direct variance estimates are calculated. Victimization summaries must be moved to the person and household files before direct variance estimation, as outlined in *Section 2.2*.

When calculating victimization totals, rates, and proportions, analysts must take care to account for series victimizations. Because each series victimization is represented as a single record on the incident file, weight adjustments are needed to account appropriately for series victimizations for all three types of estimates. These adjustments are detailed in *Sections 2.1* and *2.2*.

Modified versions of the incident-level file (for the calculation of victimization totals and proportions), the household-level file (for the calculation of property victimization rates), and the person-level file (for the calculation of personal victimization rates) have been created to facilitate the majority of NCVS analyses. The modified household- and person-level files include

the most common incident-level characteristics typically used in NCVS analyses. **Sections 2.1** and **2.2** document how these files were created. In addition, if an analyst wants to compute victimization rates for incident-level characteristics that are not included on the modified files, **Section 2.2** provides a step-by-step guide for adjusting the household- and person-level files for the calculation of victimization rates.

2.1 Preparing the Incident-Level File for Direct Variance Estimation of Totals and Proportions

Starting with the 1994 NCVS, design-consistent estimates of victimization totals and proportions can be calculated directly from a modified version of the incident file. NCVS design variables, the pseudo-stratum (*V2117*), and half-sample (*V2118*) codes are available on the household and incident files for 1994 and later years. The pseudo-stratum variable approximates the stratification in the NCVS, whereas the half-sample code approximates the primary sampling units (PSUs).

The pseudo-stratum and half-sample codes specify the NCVS design within a single year. However, NCVS samples are not independent across years, so an additional stratification variable is needed to correctly specify the overlap in the NCVS yearly samples.

- In 1996 and 2006 new samples of PSUs were phased in while the previous PSUs were phased out, so a *year group* variable must be created and added to the incident file to account for changes in the NCVS design during these times (this variable is called *YR_GRP* on the modified incident file).

- For the combined 1994 through 2015 NCVS data files, the year group variable has three levels (1994–1996, 1997–2005, and 2006–2015).[7]

NCVS samples within a single year group are not independent, as they are based on the same set of PSUs, whereas samples from different year groups are independent. To obtain the correct variance estimates for NCVS data pooled across years, software must account for the correct stratification variables (year group and pseudo-stratum) and the correct PSUs (half-samples). Examples of specifying the correct design for each software package are included in **Appendices C–F**.

---

[7] NCVS data are currently available through 2012, but the 2013–2015 NCVS data will also be part of year group 3.

Estimated totals and proportions are based on the victimization weight (*WGTVICCY*). However, to properly account for series victimizations, the victimization weight must first be adjusted to account for multiple victimizations of the same type that are represented on the incident file as a single record. A series count is calculated using variables available on the incident file. This count is capped at 10 and multiplied by the victimization weight to obtain an adjusted victimization weight. [8] This adjusted victimization weight (*newwgt*) should be used for the calculations of victimization totals and proportions. SAS, SPSS, and Stata code for creating the adjusted victimization weight is contained in *Appendix A*, and the adjusted victimization weight is available on the modified incident file.

The final step involves preparing the incident file for direct variance estimation. Because the incident file contains records only for persons and households reporting victimizations and not all sampled households, it is important to ensure that the structure of the NCVS design is appropriately captured. The incident file must contain at least one record for each NCVS PSU to calculate variances using design-consistent estimation software. PSUs are identified on the NCVS file on the basis of the year group, pseudo-stratum (*V2117*), and half-sample codes (*V2118*).

Each combination of the year group, pseudo-stratum, and half-sample codes present on the household file must also be present on the incident file. If no respondents from a particular PSU were victimized, then the entire PSU would be missing from the incident file. This requires creation of dummy records, where PSUs without reported victimizations are included on the incident file with nonzero weights and values for design variables but missing values for all incident characteristics (see example record in *Table 2-1*). In addition to adding dummy records, the analyst should add an indicator variable to the file that equals 1 if the record is a dummy record and 0 otherwise. This allows dummy records to be excluded from all calculations, as demonstrated in the examples in *Appendices C–F*.

---

[8] For more information on series counting rules and decisions, see Lauritsen, J. L., Gatewood Owens, J., Planty, M., Rand, M. R., &Truman, J. L. (2012, April). *Methods for counting high-frequency repeat victimizations in the National Crime Victimization Survey* (NCJ No. 237308). Washington, DC: U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics. Retrieved from http://www.bjs.gov/index.cfm?ty=pbdetail&iid=2240

Table 2-1.   Example Dummy Record on the Modified Incident File

DUMMY	YR_GRP	YEAR	YEARQ	V2117	V2118	IDHH	IDPER	WGTVICCY	NEWWGT	OUTCOMES
1	1	1995	1995.2	12	2	—	—	1	1	—

The year group variable, pseudo-stratum, half-sample code, series-adjusted victimization weight, and dummy records needed for accurate design-consistent variance estimation are included on the modified incident file, so when the modified file is being used no further manipulation is needed before the victimization totals and proportions are calculated directly.

2.2   Preparing the Household- and Person-Level Files for Direct Variance Estimation of Victimization Rates

Unlike victimization totals and proportions, victimization rates are derived from multiple files and multiple analysis weights. The numerator of a victimization rate is the estimated number of victimizations, which is obtained from the incident file using the victimization weight. The denominator is the estimated population size, coming either from the household file and the household weight (for property crimes) or the person file and the person weight (for personal crimes).

Currently available software capable of direct variance estimation requires a single input file and a single analysis weight, so data processing steps are required to prepare NCVS data for direct variance estimation. Modified household and person files have been created, allowing for the estimation of victimization rates of all major personal and property crime types (rape/sexual assault, robbery, simple assault, aggravated assault, personal theft, household burglary, motor vehicle theft, and theft) by several incident characteristics (reported to police, victim-offender relationship, weapon category, weapon use, and injury type). The modified household- and person-level files assume that analysts will include the series adjustment outlined in *Section 2.2.2* and will exclude victimizations occurring outside of the United States. For other incident characteristics of interest, additional data processing steps are required. Victimization summaries and victimization weights must be moved from the incident file to the household file (for property crimes) or the person file (for personal crimes). Finally, a victimization adjustment factor is calculated for use in direct variance estimation. These steps can be carried out as follows.

2.2.1 Step 1: Identify Records with Victimization Characteristics of Interest

The first step in preparing for the calculation of victimization rates is to identify the victimization characteristics of interest on the incident-level file. The analyst must create an indicator variable that equals 1 if the record is associated with the characteristics of interest and 0 otherwise. Care must be taken to include only the victimization types of interest. For example, if the analyst is interested only in victimizations occurring within the United States (i.e., *V4022* not equal to 1), then victimizations occurring outside the United States should be assigned a value of 0 for this indicator variable. If the analyst is interested only in a single level of a victimization characteristic (e.g., victimizations reported to police), then only one indicator variable needs to be created. However, if the analyst is interested in multiple levels of the victimization characteristic (e.g., both victimizations reported and not reported to police), then indicator variables must be created for each level of the victimization characteristic.

2.2.2 Step 2: Create a Victimization Summary File

After creating one or more indicator variables associated with the victimization characteristics of interest on the incident-level file, the analyst must create the victimization summary file. Victimization summaries are the counts of victimizations of a specific type by person (for personal crimes) or household (for property crimes) and reporting period. The victimization summary file will contain one variable for each indicator created in Step 1, and these variables will be equal to the count of victimizations of that type for each household (for property crimes) or person (for personal crimes) and reporting period. The victimization summary should also include the unadjusted victimization weight (*WGTVICCY*). Because the unadjusted victimization weight is the same within each person (for personal crimes) or household (for property crimes) by reporting period, it does not matter from which record the victimization weight is retained if there are multiple records per reporting period.

The unadjusted victimization weight, rather than the series-adjusted victimization weight, is included on the victimization summary file because the series adjustment is made within the victimization count. The victimization count calculated in the victimization summary is weighted by the series count to account for multiple occurrences of the same crime represented as a single

record on the file.[9] Code documenting the calculation of the series count is contained in **Appendix A**, and the series count is included on the modified person- and household-level files as the variable *serieswgt*.

For personal crimes, the victimization summary file created in Step 2 has one record for each person and reporting period contained on the incident-level file, along with the count of victimizations with the characteristics of interest (one variable for each indicator created in Step 1) and the victimization weight. For property crimes, the victimization summary file contains one record for each household by reporting period contained on the incident-level file.

2.2.3 Step 3: Merge Victimization Summary File onto the Appropriate File

Victimization summaries must then be merged onto the household file (for property crimes) or the person file (for personal crimes) using the appropriate identifiers available on the PUFs (i.e., *IDPER* and *YEARQ* for personal crimes and *IDHH* and *YEARQ* for property crimes). The victimization count is set to 0 for records on the household- or person-level file that were not contained on the victimization summary file (i.e., persons who did not experience victimizations).

2.2.4 Step 4: Calculate the Victimization Adjustment Factor

After victimization summaries have been merged onto the person file or the household file, the modified person or household file will contain two weights: (1) person or household weight and (2) victimization weight. Software currently available for the calculation of design-consistent estimation allows for the input of only a single weight, so one additional step is needed before victimization rates are calculated.

The unadjusted victimization weight (*WGTVICCY*) is based on the person weight for personal crimes and the household weight for property crimes. The following relationship exists between the weights:

$$v_{ijk} = 2 * BF_{ijk} * w_{ijk},$$

---

[9] For example, if a respondent reported eight similar robberies in a single incident record as a series, that record should be counted as eight robberies rather than one.

where v_{ijk} is the unadjusted victimization weight (*WGTVICCY*) for household i, respondent j, and reporting period k; BF_{ijk} is the bounding factor for household i, respondent j, and reporting period k; and w_{ijk} is either the person weight (*WGTPERCY*) for household i, respondent j, and reporting period k (for personal crimes) or the household weight (*WGTHHCY*) for household i and reporting period k (for property crimes).

The multiplier by two is present because of the annualization of person and household weights. Respondents are interviewed every 6 months (twice a year), so the person and household weights must be weighted down by a factor of one half on the annual file to account for the 6-month reporting periods. Victimizations are cumulative across the year, so they are not weighted down like the person and household weights. The bounding adjustment takes into account the effects of telescoping during the first interview. The telescoping phenomenon is when respondents report victimizations that occurred outside of the 6-month reference period and thus an adjustment is needed to correct for these reported victimizations. The bounding factor adjustment BF_{ijk}, where $0 < BF_{ijk} \leq 1$, weights down victimizations reported during the first interview.

Because the relationship between the victimization weight and the person and household weights is known, the victimization weight can be parsed into two components: (1) the base weight (household or person weight, depending on the type of crime) and (2) the victimization adjustment factor, which takes into account the annualization and bounding factor adjustments. The victimization adjustment factor A_{ijk} is calculated as follows:

$$A_{ijk} = \frac{v_{ijk}}{w_{ijk}} = 2 * BF_{ijk},$$

where w_{ijk} is the person weight (*WGTPERCY*) for personal crimes or the household weight (*WGTHHCY*) for household crimes. A_{ijk} is set equal to 0 if no incidents were reported by household i, respondent j in reporting period k (i.e., if the victimization weight is not defined for that person or household). After the adjustment factor is calculated, the modified household and person files are ready for direct variance estimation of victimization rates. The adjustment factor is included on the modified household- and person-level files as the variable *ADJINC_WT*.

A detailed example for setting up the person file for the calculation of a personal victimization rate is provided in **Section 2.3** below. The procedure for setting up the household file for the calculation of property victimization rates is analogous, replacing the person file with the household file and the person identifier with the household identifier. Example code for setting up both the person and household files for direct variance estimation is contained in *Example 7* of ***Appendices C–F*** for SAS, SUDAAN, SPSS, and Stata, respectively.

2.3 Example

In this example, we will set up the person file for the calculation of a personal victimization rate, the rate of violent victimizations occurring at night. This example focuses on the case where the analyst is interested only in a single level of a victimization characteristic, but the situation where the analyst is interested in multiple levels of the victimization characteristic is analogous (adding additional indicator variables in Step 1 as needed). ***Table 2-2*** contains a simplified version of the incident file, limited to a subset of personal victimizations and the subset of variables from the incident file needed for this calculation. The file contains the household and person identifiers, an indicator for the reporting period, the series count, whether or not the crime occurred within the United States, the victimization weight, and crime classifications for the type of crime and when the crime occurred. Although the household identifier is not needed in the creation of the victimization summary file, it is retained on the file in the example below for clarification purposes.

2.3.1 Step 1: Identify Records With Victimization Characteristics of Interest

The first step is to add an indicator variable to the incident file that is 1 if the incident has the victimization characteristics of interest and is 0 otherwise. Any victimizations that are to be excluded from the victimization rate calculation, such as victimizations occurring outside the United States, should have a value of 0 for the indicator variable.

Table 2-2. Example Incident File (Subset of Personal Crimes)

IDHH[1]	IDPER[2]	Year/Quarter[3]	Series Count[4]	Victimization Wgt[5]	Crime Type[6]	Time of Crime[7]	Within the U.S.[8]
1	11	1	1	v_{111}	Personal theft	Night	Yes
1	11	2	1	v_{112}	Robbery	Night	Yes
1	11	2	6	v_{112}	Simple assault	Night	Yes
1	11	2	1	v_{112}	Simple assault	Day	No
1	12	2	1	v_{122}	Aggravated assault	Unknown	Yes
2	21	1	1	v_{231}	Simple assault	Night	Yes
2	21	1	1	v_{231}	Rape/sexual assault	Night	Yes
2	21	2	8	v_{232}	Simple assault	Day	No
2	23	1	1	v_{251}	Personal theft	Night	No
2	23	2	1	v_{252}	Robbery	Night	Yes

[1] IDHH is the household identification number.
[2] IDPER is the person identification number.
[3] The year and quarter indicator represent the reporting period and are included on the incident file as *YEARQ*.
[4] The series count is equal to *serieswgt*, as outlined in **Appendix A**.
[5] The victimization weight is *WGTVICCY*. The victimization weight for household i, person j, and reporting period k is represented as v_{ijk}.
[6] Crime classifications can be made using variable *V4529* on the public use file (PUF).
[7] Time of crime classifications can be made using variable *V4021B* on the PUF.
[8] Crimes occurring within the United States can be identified using variable *V4022* on the PUF.

In this example, three incident characteristics are of interest: (1) violent victimizations,[10] (2) victimizations occurring at night, and (3) victimizations that occurred within the United States. Only victimizations with all three characteristics receive a value of 1. The indicator variable is shown in **Table 2-3** in the "violent and night" column.

---

[10] Violent victimizations include rape/sexual assault, robbery, aggravated assault, and simple assault.

Table 2-3. Example Modified Incident File

IDHH[1]	IDPER[2]	Year/ Quarter[3]	Series Count[4]	Victimization Wgt[5]	Crime Type[6]	Time of Crime[7]	Within the U.S.[8]	Violent and Night
1	11	1	1	v_{111}	Personal theft	Night	Yes	0
1	11	2	1	v_{112}	Robbery	Night	Yes	1
1	11	2	6	v_{112}	Simple assault	Night	Yes	1
1	11	2	1	v_{112}	Simple assault	Day	No	0
1	12	2	1	v_{122}	Aggravated assault	Unknown	Yes	0
2	21	1	1	v_{231}	Simple assault	Night	Yes	1
2	21	1	1	v_{231}	Rape/sexual assault	Night	Yes	1
2	21	2	8	v_{232}	Simple assault	Day	No	0
2	23	1	1	v_{251}	Personal theft	Night	No	0
2	23	2	1	v_{252}	Robbery	Night	Yes	1

[1] IDHH is the household identification number.
[2] IDPER is the person identification number.
[3] The year and quarter indicator represent the reporting period and are included on the incident file as *YEARQ*.
[4] The series count is equal to *serieswgt*, as outlined in ***Appendix A.***
[5] The victimization weight is *WGTVICCY*. The victimization weight for household *i*, person *j*, and reporting period *k* is represented as v_{ijk}.
[6] Crime classifications can be made using variable *V4529* on the public use file (PUF).
[7] Time of crime classifications can be made using variable *V4021B* on the PUF.
[8] Crimes occurring within the United States can be identified using variable *V4022* on the PUF.

2.3.2 Step 2: Create Victimization Summary File

After victimizations are identified on the basis of the characteristics of interest—in this case, violent victimizations occurring at night in the United States—the total number of victimizations of interest for each person and reporting period is calculated, weighted by the series count. This person-by-reference period file, the victimization summary file, is presented in *Table 2-4*. Note that for persons who did not experience any violent victimizations at night during a reporting period, the count is set equal to 0. As shown in Table 2-2, person 11 in reporting period 2 had two records for violent victimizations at night. One was a robbery for which the respondent could recall the details of the crime (so it has a series count of 1). The other was a series of six simple assaults for which the respondent could not recall the specifics of each incident. There are a total of seven violent victimizations at night for this respondent and reporting period: one non-series incident and one series incident with a series count of 6. The victimization summary file should include the household and person identifiers along with the reporting period, victimization weight, and the victimization count for the victimization characteristics of interest (one variable for each indicator identified in Step 1).

Table 2-4. Example Victimization Summary File

IDHH[1]	IDPER[2]	Year/Quarter[3]	Victimization Wgt[4]	Victimization Count (Ct)[5]
1	11	1	v_{111}	0
1	11	2	v_{112}	7
1	12	2	v_{122}	0
2	21	1	v_{231}	2
2	21	2	v_{232}	0
2	23	1	v_{251}	0
2	23	2	v_{252}	1

[1] IDHH is the household identification number.
[2] IDPER is the person identification number.
[3] The year and quarter indicator represent the reporting period and are included on the incident file as *YEARQ*.
[4] The victimization weight is *WGTVICCY*. The victimization weight for household *i*, person *j*, and reporting period *k* is represented as v_{ijk}.
[5] The victimization count (Ct) is the total number of violent victimizations occurring at night, per person and reporting period.

2.3.3   Step 3: Merge Victimization Summary File Onto the Person-Level File

The victimization summary file is then merged onto the person-level file by the person identification number (*IDPER*) and year and quarter indicator (*YEARQ*), creating the modified person file in ***Table 2-5***. For persons and reporting periods that are not on the victimization summary file because they did not have any reported victimizations, the victimization count is set equal to 0. These records will also have missing values for the victimization weight, as the victimization weight is defined only for reported victimizations. The right two columns (the victimization weight and the victimization count) come from the victimization summary file in Table 2-4, and the remaining columns are from the original person file. If the victimization characteristic of interest had more than one level (e.g., if the analyst was interested in violent victimizations occurring both during the day and at night), then multiple victimization counts would be added to the modified person file—or, one for each level of the victimization characteristics of interest.

Table 2-5. Example of Modified Person File

IDHH[1]	IDPER[2]	Year/ Quarter[3]	Person Weight[4]	Victimization Wgt[5]	Victimization Count (Ct)[6]
1	11	1	p_{111}	v_{111}	0
1	11	2	p_{112}	v_{112}	7
1	12	1	p_{121}		**0**
1	12	2	p_{122}	v_{122}	0
2	21	1	p_{231}	v_{231}	2
2	21	2	p_{232}	v_{232}	0
2	22	1	p_{241}		**0**
2	22	2	p_{242}		**0**
2	23	1	p_{251}	v_{251}	0
2	23	2	p_{252}	v_{252}	1

Persons/reporting periods not on victimization summary file are set equal to 0

[1] IDHH is the household identification number.
[2] IDPER is the person identification number.
[3] The year and quarter indicator represents the reporting period and is included on the incident file as *YEARQ*.
[4] The person weight is *WGTPERCY*. The person weight for household i, person j, and reporting period k is represented as p_{ijk}.
[5] The victimization weight is *WGTVICCY*. The victimization weight for household i, person j, and reporting period k is represented as v_{ijk}.
[6] The victimization count (Ct) is the total number of violent victimizations occurring at night, per person and reporting period.

2.3.4 Step 4: Calculate the Victimization Adjustment Factor

The final step of preparing the modified person file for direct variance estimation is to calculate the victimization adjustment factor. The victimization adjustment factor is the unadjusted victimization weight divided by the person weight (v_{ijk}/w_{ijk}). As discussed previously, this factor is equal to 2 times the bounding factor adjustment. For persons and reference periods for which no victimizations were reported, the victimization adjustment factor is set equal to 0. The final modified person file is presented in *Table 2-6*.

18

Table 2-6.   Example Final Modified Person File

IDHH[1]	IDPER[2]	Year/ Quarter[3]	Person Weight[4]	Victimization Wgt[5]	Victimization Count (Ct)[6]	Victimization Adjustment Factor (A)[7]
1	11	1	p_{111}	v_{111}	0	v_{111} / p_{111}
1	11	2	p_{112}	v_{112}	7	v_{112} / p_{112}
1	12	1	p_{121}		0	0
1	12	2	p_{122}	v_{122}	0	v_{122} / p_{122}
2	21	1	p_{231}	v_{231}	2	v_{231} / p_{231}
2	21	2	p_{232}	v_{232}	0	v_{232} / p_{232}
2	22	1	p_{241}		0	0
2	22	2	p_{242}		0	0
2	23	1	p_{251}	v_{251}	0	v_{251} / p_{251}
2	23	2	p_{252}	v_{252}	1	v_{252} / p_{252}

[1] IDHH is the household identification number.

[2] IDPER is the person identification number.

[3] The year and quarter indicator represents the reporting period and is included on the incident file as *YEARQ*.

[4] The person weight is *WGTPERCY*. The person weight for household *i*, person *j*, and reporting period *k* is represented as p_{ijk}.

[5] The victimization weight is *WGTVICCY*. The victimization weight for household *i*, person *j*, and reporting period *k* is represented as v_{ijk}.

[6] The victimization count (Ct) is the total number of violent victimizations occurring at night, per person and reporting period.

[7] The victimization adjustment factor (A) is equal to the victimization weight divided by the person weight. This factor is called *ADJINC_WT* on the modified person and household files.

SECTION 3. OBTAINING DIRECT VARIANCE ESTIMATES

Once modified incident-, person-, and household-level files are derived (as outlined in *Section 2*), or using the newly available modified files, estimated victimization totals, proportions, and rates can be computed directly for single or pooled-year NCVS estimates using TSL. TSL approximates design-consistent variances of linear and nonlinear functions under the specified sample design on the basis of linear approximations.

When making any type of NCVS estimate, it is critical that the input file is not subset (i.e., that cases are not removed from the file) beyond limiting the file on the basis of year. This applies to the modified household, person, and incident files. For example, if an analysis will be limited to a specific domain (e.g., crimes occurring within the United States or crimes committed against females or against low-income households), observations outside of the domain of interest should not be deleted from the file. Direct variance estimation requires that the NCVS design be represented on the input file, and deleting observations risks removing PSUs from the file. Removing PSUs from the file makes the assumption that the population of interest (e.g., crimes with particular characteristics) did not occur among any households or persons within the removed PSU, including those not in the NCVS sample.

Rather than deleting records from the input file, use statements within the software to limit the analysis to the domain of interest. Examples in each software package are included in *Appendices C–F*. Because all PSUs are represented on the three input files for each year of NCVS data, the files can be subset to a specific year of interest to reduce the file size for single-year analyses.

3.1   Estimate Types

The three primary NCVS estimate types are victimization totals, proportions, and rates. *Section 3.1.1* outlines the calculation of victimization totals, *Section 3.1.2* outlines the calculation of victimization proportions, and *Section 3.1.3* outlines the calculation of victimization rates.

3.1.1 Victimization Totals

Victimization totals are the number of criminal victimizations with a given characteristic (e.g., the estimated number of simple assaults committed against females; the estimated number of household burglaries in the Northeast census region). Victimization totals are calculated from the modified incident file using the series-adjusted victimization weight (*newwgt*).[11] The estimated victimization total for domain D, $\hat{t}_D$, is

$$\hat{t}_D = \sum_{ijkl \in D} v_{ijkl},$$

where v_{ijkl} is the series-adjusted victimization weight (*newwgt*) for household i, respondent j, reporting period k, and victimization l. The series-adjusted victimization weights are summed for victimizations l with the domain characteristics of interest D for all households i, respondents j, and reporting periods k present on the modified incident file.

Analysis domains can be defined on the basis of crime types (e.g., violent crimes, property crimes), characteristics of victims (e.g., age, sex, household income), or characteristics of the victimizations (e.g., victimizations reported to police, victimizations committed with a weapon present). Domains could also be a combination of all of these types of characteristics. For example, an analyst might want to estimate the total number of violent victimizations that were committed against males and were reported to police. Regardless of the software package used, the first steps in calculating a victimization total are to properly identify the domain of interest and to flag crimes that are within the specified domain with an indicator variable (*IND*=1 if the crime is within domain *D*, *IND*=0 otherwise). *Examples 1* and *2* in ***Appendices C–F*** provide examples for calculating victimization totals for single and aggregated years in SAS, SUDAAN, SPSS, and Stata, respectively.

3.1.2 Victimization Proportions

Victimization proportions provide characteristics of victimizations or victims (e.g., the estimated percentage of robberies reported to the police, the estimated distribution of property

---

[11] If series victimizations are excluded from estimates, the unadjusted victimization weight (*WGTVICCY*) can be used in the calculation of totals. This applies to the calculation of victimization proportions as well (see ***Section 3.1.2***). Analysts should clearly document whether reported estimates include or exclude series victimizations.

crime victims by household income). Victimization proportions are calculated from the modified incident file using the series-adjusted victimization weight (*newwgt*). The estimated victimization proportion for domain D across level a of covariate A, $\hat{p}_{A_a,D}$, is

$$\hat{p}_{A_a,D} = \frac{\displaystyle\sum_{ijkl \in A_a, D} v_{ijkl}}{\displaystyle\sum_{ijkl \in D} v_{ijkl}},$$

where v_{ijkl} is the series-adjusted victimization weight (*newwgt*) for household i, respondent j, reporting period k, and victimization l.

In the numerator, series-adjusted victimization weights are summed for victimizations l with the domain characteristics of interest D associated with level a of covariate A for all households i, respondents j, and reporting periods k present on the modified incident file. In the denominator, victimization weights are summed for victimizations l with the domain characteristics of interest D for all households i, respondents j, and reporting periods k present on the modified incident file.

Similar to victimization totals, domains can be defined on the basis of crime types, characteristics of victims, characteristics of victimizations, or any combination of these. The first step in calculating a proportion is to define the domain of interest and to flag victimizations within this domain with an indicator variable (*IND*=1 if the crime is within domain D, *IND*=0 otherwise). For proportions, the analyst must also specify the covariate A for which the distribution of victimizations within domain D will be calculated. For example, an analyst might want to calculate the distribution of motor vehicle thefts in urban areas by household income. In this example, motor vehicle thefts in urban areas are the domain characteristics of interest while household income is the covariate. It is also possible to specify multiple covariates (e.g., the distribution of violent victimizations by type of crime and weapon category). *Examples 3* and *4* in *Appendices C–F* demonstrate how to calculate victimization proportions for single and aggregated years in SAS, SUDAAN, SPSS, and Stata, respectively. *Example 4* specifies multiple covariates of interest.

3.1.3   Victimization Rates

Victimization rates are the estimated number of victimizations per 1,000 persons or households in the population (e.g., the rate of simple assaults against females per 1,000 women, the rate of thefts reported to the police per 1,000 households). The estimated victimization rate for crime C in domain D, $V\hat{R}_{C,D}$, is

$$V\hat{R}_{C,D} = \frac{\sum_{ijkl \in C,D} v_{ijkl}}{\sum_{ijk \in D} w_{ijk}} * 1,000,$$

where v_{ijkl} is the series-adjusted victimization weight (*newwgt*) for household i, respondent j, reporting period k, and victimization l; and w_{ijk} is the person weight (*WGTPERCY*) for household i, respondent j, and reporting period k (if crime C is a personal crime) or the household weight (*WGTHHCY*) for household i and reporting period k (if crime C is a property crime).

For direct variance estimation, this formula can be rewritten to avoid multiple input files and weights. The victimization weight can be parsed into the base weight (the household or person weight) and the victimization adjustment factor (*ADJINC_WT*), as discussed in *Section 2.3*. With the revised formulation, the victimization rates for property crimes are calculated from the modified household-level file using the household weight (*WGTHHCY*). Personal victimization rates are calculated from the modified person-level file using the person weight (*WGTPERCY*). The estimated victimization rate for crime C in domain D, $V\hat{R}_{C,D}$, is

$$V\hat{R}_{C,D} = \frac{\sum_{ijk \in D} w_{ijk} * A_{ijk} * n_{ijk,C}}{\sum_{ijk \in D} w_{ijk}} * 1,000,$$

where w_{ijk} is the person weight (*WGTPERCY*) for household i, respondent j, and reporting period k (if crime C is a personal crime) or the household weight (*WGTHHCY*) for household i and reporting period k (if crime C is a property crime); A_{ijk} is the victimization adjustment factor (*ADJINC_WT*) for household i, respondent j, and reporting period k; and $n_{ijk,C}$ is the series-adjusted victimization count—that is, the number of crime C victimizations associated with

household i, respondent j, and reporting period k. The value $n_{ijk,C}$ is calculated from the incident-level file as follows:

$$n_{ijk,C} = \sum_{ijk \in D} s_{ijkl} * I_{ijkl},$$

where s_{ijkl} is the series count for household i, respondent j, reporting period k (*serieswgt*), and victimization l and I_{ijkl} is an indicator variable that equals 1 if victimization l is associated with crime C and 0 otherwise.

The domain D is defined on the basis of characteristics of victims (e.g., age, sex, household income, census region), and the crime type C is defined on the basis of characteristics of victimizations (e.g., reported to police, weapon use, victim-offender relationship). *Examples 5 and 6* in ***Appendices C–F*** show how to calculate victimization rates for single and aggregated years on the basis of victimization characteristics already included on the modified household and person level files in SAS, SUDAAN, SPSS, and Stata, respectively. *Example 7* demonstrates how to calculate victimization rates for victimization characteristics not included on the file (i.e., how to create victimization summaries and move them onto the household and person files before the calculation of victimization rates).

3.2   Computing Direct Variance Estimates

SAS, SUDAAN, SPSS, and Stata are four software packages capable of design-consistent variance estimation. ***Appendices C–F*** contain examples of how to estimate victimization totals, proportions, and rates for single- and pooled-year estimates in these four software packages. Also included are examples of identifying low-quality estimates as outlined in ***Section 4***. Regardless of the software package used, analyses of NCVS data must specify the sample design and the appropriate analysis weight.

For all estimate types, the input file can be either a single year of NCVS data or stacked files including multiple years of data. Beyond subsetting by year, no records should be deleted from the files before analysis. This ensures that the software appropriately estimates variances on the basis of the NCVS design. All four software packages contain methods for limiting the analysis to a specified set of cases while maintaining correct variance estimates: the *domain* statement in SAS, the *subpopn* statement in SUDAAN, and the *subpop* statements in SPSS and

Stata). *Table 3-1* provides variable names for key design elements and procedure specifications for each software package for all three types of NCVS estimates.

Table 3-1.   NCVS Design Elements and Procedure Names by Estimate Type

Design Element	Totals & Proportions	Personal Victimization Rates	Property Victimization Rates
Input file	adjusted incident	adjusted person	adjusted household
Stratification variables	yr_group, v2117	yr_group, v2117	yr_group, v2117
Replicate/primary sampling unit	v2118	v2118	v2118
Analysis weight	newwgt	wgtpercy	wgthhcy
Adjustment factor	n/a	adjinc_wt	adjinc_wt
SAS procedures	surveymeans/surveyfreq	surveymeans	surveymeans
SUDAAN procedures	descript/crosstab	vargen/descript	vargen/descript
SPSS procedures	csdescriptives/cstabulate	csdescriptives	csdescriptives
Stata procedures	total/prop	mean	mean

SECTION 4. RECOMMENDATIONS ON IDENTIFYING LOW-QUALITY ESTIMATES

Before reporting NCVS estimates, analysts should consider whether or not the estimates are stable enough to provide reliable conclusions. In general terms, an estimate should not vary widely because of sampling variation, nor should it be overly influenced by a small number of observations. When neither of these situations occur, an estimate is said to be reliable because the users can have confidence that conclusions drawn from the estimate are likely not due to random chance. Unreliable estimates are difficult to interpret because the observed values may be due to random chance or may not represent the target population well. This section provides recommendations for flagging NCVS estimates with low reliability and a general rounding rule for low-prevalence estimates. More details about the reliability rules can be found in the direct variance technical report.[12]

Reliability flag recommendations are based on the unweighted sample size and the percent relative standard errors (RSEs) of the estimate. If an analyst is accounting for series victimizations, the unweighted sample size used for flagging unreliable estimates should be the series-adjusted count of victimizations. For example, a series victimization with a series count of seven would count as seven victimizations in the unweighted sample size, whereas a non-series victimization would count as only one. The percent RSE of an estimate x is computed by taking the ratio of the estimated standard error ($SE(x)$) to the estimate (x). This ratio is then expressed as a percentage:

$$\% \, RSE(x) = \frac{SE(x)}{x} * 100.$$

As outlined in *Sections 4.1–4.3*, it is recommended that estimates with a percent RSE greater than 30 or unweighted sample sizes of less than or equal to 10 be flagged as unreliable. As further discussed in the direct variance technical report, these recommendations continue several features that have been used by the NCVS for many years while tightening the RSE criteria to be consistent with those of many other major national surveys. Increasing the

---

[12] Williams et al., ibid.

unweighted sample size threshold to other values that were considered (≤ 20 or ≤ 30) would lead to an unacceptable number of estimates identified as unreliable.

In addition to flagging low-quality estimates, it is recommended that any estimate that rounds to 0, regardless of the estimate type or the number of significant digits presented, not be displayed. For example, if an analyst is reporting a victimization rate to one decimal place and the estimated rate is 0.045, the estimate would round to 0.0 and should not be presented. In BJS reports, estimates that round to 0 are suppressed with a dash symbol.

Section 4.1 provides recommendations for identifying unreliable victimization totals, *Section 4.2* for victimization proportions, and *Section 4.3* for victimization rates. *Appendices C–F* provide examples of implementing the flagging and rounding rules for each estimate type.

4.1   Victimization Totals

For victimization totals, it is recommended that estimates meeting the following criteria be flagged as unreliable. The sample size for a victimization total is the series-adjusted count of victimizations used to calculate the estimate:

- RSE > 30% or

- count sample size ≤ 10.

4.2   Victimization Proportions

The methodology used to flag unreliable victimization proportions (which can be expressed as percentages) is different from that used for victimization totals and rates. When a percentage (P) is near 0%, the RSE will become large and can understate the precision of the estimated percentage. In addition, it is desirable that a reliability criterion for a percentage similarly treat the situations where P approaches either 0% or 100%, because the variance of P and $100-P$ are the same; however, the RSEs for P and $100-P$ are not the same. For this reason, reliability flagging is based on the RSE of $\log(P)$ or $\log(100-P)$ rather than the RSE of P. The log transformation lessens the impact of P being near 0%, whereas the switch between P and $100-P$ makes the rule symmetric.

For victimization proportions or percentages, it is recommended that estimates meeting the following criteria be flagged as unreliable. The denominator sample size is the series-adjusted count of victimizations in the denominator of the proportion or percentage:

- RSE > 30%, where

    - if the percentage is ≤ 50%, use the RSE of log(P) or
    - if the percentage is > 50%, use the RSE of log(100−P); or
- denominator sample size ≤ 10.

4.3   Victimization Rates

For victimization rates, it is recommended that estimates meeting the following criteria be flagged as unreliable. The numerator sample size for a victimization rate is the series-adjusted count of victimizations in the numerator of the rate:

- RSE > 30% or

- numerator sample size ≤ 10.

APPENDIX A
CALCULATING THE SERIES ADJUSTMENT

As discussed in *Section 2*, when calculating totals, proportions, and rates, it is important to account for *series victimizations*. Series victimizations occur when an NCVS respondent recalls at least six criminal incidents of a similar nature but cannot recall the dates and other details of the individual incidents well enough to report them separately. Each series victimization is represented as a single record on the incident file, and the number of incidents linked to that series is recorded. Weight adjustments are needed to appropriately account for series victimizations for all three types of estimates.

When series victimizations are included, calculations of totals and proportions should be based on an adjusted victimization weight (*newwgt*), as defined in the example code below. For the calculation of victimization rates, total counts of victimizations must be weighted by the series weight (*serieswgt*) before moving victimization summaries onto the household- or person-level files. The SAS, SPSS, and Stata code below should be applied to the NCVS incident file. This code creates the adjusted victimization weight and the series weight.

A.1   SAS Syntax

```
**create series weight;

if (v4017 in (1,8)) or (v4018 in (2,8)) or (v4019 in (1,8)) then
series=1; else series=2;

serieswgt=1;

**create temporary variable for alternative weight;

if v4016 not in (997 998) then n10v4016=v4016;
if n10v4016>10 then n10v4016=10; if (series=2) then
serieswgt=n10v4016;
if (series=2 and n10v4016=.) then serieswgt=6;

newwgt=wgtviccy*serieswgt;
```

A.2   SPSS Syntax

```
***create series weight.

missing values all().

do  if (any(v4017,1,8) or
     any(v4018,2,8) or
     any(v4019,1,8)).
+  compute series=1.
```

```
else.
+  compute series=2.
end if.

variable label
 series     'incident frequency'.

value label
 series  1 'not a series crime'
      2 'series crime'/.

execute.

compute serieswgt=1.
missing values v4016 (997 998).
compute n10v4016=v4016.

**create temporary variable for alternative weight.

recode n10v4016 (10 thru hi=10).

if (series=2) serieswgt=n10v4016.
if (series=2 and missing(n10v4016)) serieswgt=6.
compute newwgt=wgtviccy*serieswgt.

execute.
```

A.3   Stata Syntax

```
/*create series weight*/
generate series=2
replace series=1 if inlist(v4017,1,8) | inlist(v4018,2,8) |
inlist(v4019,1,8)

**create temporary variable for alternative weight;

generate n10v4016=v4016 if inlist(v4016,997,998)==0
replace n10v4016=10 if n10v4016 >10

generate serieswgt=1
replace serieswgt=n10v4016 if series==2
replace serieswgt=6 if series==2 & n10v4016==.

generate newwgt=wgtviccy*serieswgt
label variable series "incident frequency"
label define series
label values series series
label define series 1"not a series crime" 2"series crime
```

APPENDIX B
GENERALIZED VARIANCE FUNCTION EQUATIONS AND PARAMETERS: 1993–2012

The GVF equations for victimization totals, percentages, and rates are presented in **Sections B.1**, **B.2**, and **B.3**, respectively. **Section B.4** contains the GVF parameters that have traditionally been used to calculate standard errors for the NCVS. These parameters, which vary by year and estimate type, should be used for collection-year estimates that include series victimizations.

B.1 GVF Equation for Victimization Totals

$$G_t(\hat{z}; a, b, c) = a\hat{z}^2 + b\hat{z} + c\hat{z}\sqrt{\hat{z}}$$

where a, b, and c are the GVF parameters contained in **Section B.4** and $\hat{z}$ is the estimated victimization total.

B.2 GVF Equation for Victimization Percentages

$$G_p(\hat{y}, \widehat{N}_p; b, c) = b\frac{\hat{y}(100 - \hat{y})}{\widehat{N}_p} + c\frac{\hat{y}(\sqrt{100\hat{y}} - \hat{y})}{\sqrt{\widehat{N}_p}}$$

where b and c are the GVF parameters contained in **Section B.4**, $\hat{y}$, is the estimated victimization percent, and $\widehat{N}_p$ is an estimate of the total size of the subpopulation of interest.

B.3 GVF Equation for Victimization Rates

$$G_r(\hat{x}, \widehat{N}_p; b, c) = b\frac{\hat{x}(1000 - \hat{x})}{\widehat{N}_p} + c\frac{\hat{x}\sqrt{1000\hat{x}} - \hat{x}}{\sqrt{\widehat{N}_p}}$$

where b and c are the GVF parameters contained in **Section B.4**, $\hat{x}$, is the estimated victimization rate, and $\widehat{N}_p$ is an estimate of the total size of the subpopulation of interest.

B.4   GVF Parameters

1993 Parameter Sets[1]

	Collection Year		
	a	**b**	**c**
1. Overall person crime estimates	-0.00042893	6,085	5.818
2. Person crime domain estimates	-0.00073642	3,923	9.934
3. Overall property crime estimates	-0.00034948	838	3.406
4. Property crime domain estimates	-0.00033613	2,011	3.019

[1] The 1993 data originated from two half-samples, NCS and NCVS. The GVF calculations for 1993 are based only on the NCVS half-sample.

1994 Parameter Sets

	Collection Year		
	a	**b**	**c**
1. Overall person crime estimates	-0.00040854	2,864	5.768
2. Person crime domain estimates	-0.00051953	2,838	7.021
3. Overall property crime estimates	-0.00038080	319	3.787
4. Property crime domain estimates	-0.00033372	2,044	3.004

1995 Parameter Sets

	Collection Year		
	a	**b**	**c**
1. Overall person crime estimates	-0.00037079	2,826	5.245
2. Person crime domain estimates	-0.00058524	2,084	7.992
3. Overall property crime estimates	-0.00035466	332	3.540
4. Property crime domain estimates	-0.00033252	2,002	3.004

1996 Parameter Sets

	Collection Year		
	a	**b**	**c**
1. Overall person crime estimates	-0.00054329	1,224	7.924
2. Person crime domain estimates	-0.00071706	2,141	9.853
3. Overall property crime estimates	-0.00022338	2,041	2.062
4. Property crime domain estimates	-0.00025538	2,528	2.216

1997 Parameter Sets

	Collection Year		
	a	**b**	**c**
1. Overall person crime estimates	-0.00060876	3,973	8.758
2. Person crime domain estimates	-0.00084400	3,084	11.606
3. Overall property crime estimates	-0.00050902	-526	5.242
4. Property crime domain estimates	-0.00042269	1,609	3.955

1998 Parameter Sets   Collection Year

	a	b	c
1. Overall person crime estimates	-0.00069365	716	10.284
2. Person crime domain estimates	-0.00083876	1,593	11.674
3. Overall property crime estimates	-0.00045078	1,647	4.466
4. Property crime domain estimates	-0.00039879	2,640	3.639

1999 Parameter Sets   Collection Year

	a	b	c
1. Overall Person Crime Estimates	-0.00062265	2,974	9.132
2. Person Crime Domain Estimates	-0.00085331	2,881	11.837
3. Overall Property Crime Estimates	-0.00041482	960	4.201
4. Property Crime Domain Estimates	-0.00036800	2,455	3.383

2000 Parameter Sets   Collection Year

	a	b	c
1. Overall Person Crime Estimates	-0.00063445	1,650	9.445
2. Person Crime Domain Estimates	-0.00091124	2,138	12.745
3. Overall Property Crime Estimates	-0.00073678	-1,921	7.854
4. Property Crime Domain Estimates	-0.00061236	827	5.981

2001 Parameter Sets   Collection Year

	a	b	c
1. Overall Person Crime Estimates	-0.00052825	1,928	7.870
2. Person Crime Domain Estimates	-0.00069129	2,774	9.633
3. Overall Property Crime Estimates	-0.00059233	-793	6.276
4. Property Crime Domain Estimates	-0.00052248	1,263	5.070

2002 Parameter Sets   Collection Year

	a	b	c
1. Overall Person Crime Estimates	-0.00085677	2,087	12.901
2. Person Crime Domain Estimates	-0.00108486	1,822	15.331
3. Overall Property Crime Estimates	-0.00040390	1,053	4.142
4. Property Crime Domain Estimates	-0.00037459	2,209	3.507

2003 Parameter Sets   Collection Year

	a	b	c
1. Overall Person Crime Estimates	-0.00051622	3,646	7.750
2. Person Crime Domain Estimates	-0.00085819	2,593	12.280
3. Overall Property Crime Estimates	-0.00048584	-197	5.209
4. Property Crime Domain Estimates	-0.00047387	1,566	4.621

2004 Parameter Sets

	a	b	c
1. Overall Person Crime Estimates	-0.00035852	4,395	5.291
2. Person Crime Domain Estimates	-0.00060837	3,104	8.659
3. Overall Property Crime Estimates	-0.00033261	2,145	3.379
4. Property Crime Domain Estimates	-0.00031991	2,552	2.995

2005 Parameter Sets   **Collection Year**

	a	b	c
1. Overall Person Crime Estimates	-0.00060311	2,805	9.251
2. Person Crime Domain Estimates	-0.00081369	3,213	11.713
3. Overall Property Crime Estimates	-0.00036754	2,039	3.789
4. Property Crime Domain Estimates	-0.00036501	2,698	3.460

2006 Parameter Sets   **Collection Year**

	a	b	c
1. Overall Person Crime Estimates	-0.00056113	5,459	8.476
2. Person Crime Domain Estimates	-0.00073569	3,217	10.626
3. Overall Property Crime Estimates	-0.00025057	3,500	2.398
4. Property Crime Domain Estimates	-0.00023644	3,800	2.052

2007 Parameter Sets   **Collection Year**

	a	b	c
1. Overall Person Crime Estimates	-0.00059713	1,795	9.335
2. Person Crime Domain Estimates	-0.00107591	264	15.939
3. Overall Property Crime Estimates	-0.00027740	2,364	2.816
4. Property Crime Domain Estimates	-0.00028978	2,565	2.732

2008 Parameter Sets   **Collection Year**

	a	b	c
1. Overall Person Crime Estimates	-0.00063195	5,004	9.722
2. Person Crime Domain Estimates	-0.00076651	5,025	11.857
3. Overall Property Crime Estimates	-0.00035885	1,887	3.778
4. Property Crime Domain Estimates	-0.00032519	3,348	3.275

2009 Parameter Sets   **Collection Year**

	a	b	c
1. Overall Person Crime Estimates	-0.00054111	2,410	8.475
2. Person Crime Domain Estimates	-0.00106740	3,529	16.794
3. Overall Property Crime Estimates	-0.00030998	2,160	3.233
4. Property Crime Domain Estimates	-0.00026607	3,098	2.663

Collection Year

2010 Parameter Sets

	a	b	c
1. Overall Person Crime Estimates	-0.00086145	1,384	13.696
2. Person Crime Domain Estimates	-0.00108760	1,960	17.278
3. Overall Property Crime Estimates	-0.00020353	2,450	2.035
4. Property Crime Domain Estimates	-0.00017899	3,297	1.687

2011 Parameter Sets | | **Collection Year** | |

	a	b	c
1. Overall Person Crime Estimates	-0.00060211	2,439	9.511
2. Person Crime Domain Estimates	-0.00081383	2,309	12.916
3. Overall Property Crime Estimates	-0.00028723	4,182	2.809
4. Property Crime Domain Estimates	-0.00037148	2,981	3.852

2012 Parameter Sets | | **Collection Year** | |

	a	b	c
1. Overall Person Crime Estimates	-0.00025343	6,439	3.704
2. Person Crime Domain Estimates	-0.00054187	3,893	8.530
3. Overall Property Crime Estimates	-0.00020722	2,345	2.116
4. Property Crime Domain Estimates	-0.00020565	2,810	2.057

APPENDICES C–F

Appendices C–F contain examples of how to estimate victimization totals, proportions, and rates for single- and pooled-year estimates in these four software packages. Also included are examples of identifying low-quality estimates as outlined in Section 4. Each appendix is available as a separate document.

Appendix C   Examples in SAS

Appendix D   Examples in SUDAAN

Appendix E   Examples in SPSS

Appendix F   Examples in Stata