Bureau of Justice Statistics Research and Development Series

ASSESSING THE COVERAGE AND RELIABILITY OF SUBNATIONAL GEOGRAPHIC IDENTIFIERS IN THE NCVS PUBLIC-USE FILE

Bonnie E. Shook-Sa, MAS
Philip Lee, BS
Marcus Berzofsky, DrPH
RTI International

NCJ 249467

R&DP-2015: 03
NCJ 249467
December 28, 2015

This document was prepared using federal funds provided by the U.S. Department of Justice under cooperative agreement number 2011-NV-CX-K068. The BJS project managers were Lynn Langton, BJS Statistician, and Michael Planty, Victimization Unit Chief.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

SECTION 1. BACKGROUND

The National Crime Victimization Survey (NCVS), sponsored by the Bureau of Justice Statistics (BJS), estimates the incidence and describes the characteristics of criminal victimization in the United States. The NCVS has been a rich source of information about criminal victimization at the national level since its inception, but subnational estimates would be useful in better understanding local crime patterns and trends. As part of its subnational estimation program, BJS intends to develop generic area typologies based on various geographic, social, economic, and demographic characteristics. These generic areas will then represent all places that are similar to each other based on the characteristics of interest.

The primary objective of this analysis was to assess the coverage and reliability of the NCVS sample in the subnational geographic areas that can be created from the public-use files (PUFs): Census region, population size, and urbanicity. Ideally, BJS would like to create generic areas based on these three variables and examine patterns and trends in victimization rates across these areas. Before releasing generic area estimates, it was necessary to first examine the coverage and reliability of the samples in these areas, to determine the most appropriate survey weights to use in estimation, and to assess the best variance estimation method for producing reliable estimates.

The evaluation was broken into three parts: (1) assessment of coverage, (2) assessment of reweighting methods, and (3) assessment of the reliability of the generalized variance function (GVF) estimates for generic areas. ***Section 2*** describes the assessment of coverage within generic areas formed on the basis of two- and three-variable crosses of Census region, population size, and urbanicity. ***Section 3*** discusses the approach for reweighting the NCVS sample within generic areas and the analysis of the effects of reweighting on key NCVS estimates and precision. It also discusses recommendations on the collapsing of levels within generic areas and pooling of data across multiple years to achieve adequate precision. Because GVFs have traditionally been used for NCVS estimation but were not designed for use in subnational areas, ***Section 4*** assesses the reliability of GVF estimates within generic areas. ***Section 5*** summarizes

key findings across the analysis and provides overall recommendations for calculating estimates within the generic areas formed with these three subnational variables.

SECTION 2. ASSESSING THE COVERAGE OF THE NCVS SAMPLE IN GENERIC AREAS DEFINED FROM THE PUBLIC-USE FILE

2.1 Overview of Coverage Assessment

As proposed by Planty (2012), the three geographic identifiers available on the NCVS PUF (Census region, population size, and urbanicity) can be crossed to produce four types of generic areas:

- Census region × population size,
- Census region × urbanicity,
- Population size × urbanicity, and
- Census region × population size × urbanicity.

With appropriate collapsing of levels and pooling of data, the sample sizes within these generic areas are sufficient to produce estimates with reasonable precision (as further discussed in ***Section 3.4***). However, the NCVS was designed to produce exclusively national estimates, which means analysis weights were created to produce representative victimization counts, rates, and proportions only at the national level without regard for smaller geographic areas. Thus, subnational estimates could exhibit systematic bias from undercoverage due to variation in the primary sampling units selected within these areas and how the sample was weighted (i.e., nonresponse and poststratification adjustments do not control weights at the subnational level). The first step in evaluating the reliability of estimates in the four types of generic areas was to assess the coverage of the NCVS sample within these areas by comparing weighted distributions of key person- and household-level demographic characteristics to external gold-standard control totals. This section documents how gold-standard estimates were derived and includes the results of the coverage assessment.

2.2 Defining Comparable Areas

To assess the coverage of the NCVS sample within generic areas, gold-standard population estimates were identified for comparison to NCVS population estimates. The two

primary sources of local population estimates in the United States are the decennial Census and the American Community Survey (ACS). Decennial Censuses before 2010 included both a short form questionnaire, which was administered to all households and contained basic demographic information, and a long form questionnaire, which was administered to one in six households and provided more detailed socioeconomic information about the population. In 2010, only the short form was administered to all households, and additional data previously obtained from the long form questionnaire are now provided by the ACS. The ACS began in 2005 and produces 1-, 3-, and 5-year national and local population estimates.

Both the decennial Census and the ACS provide estimates that can be used to assess the coverage of persons and households in generic areas defined by Census region, population size, and urbanicity. However, the local area estimates needed to make population size and urbanicity classifications are available in the ACS only for 5-year data, limiting the utility of publicly available ACS data for this evaluation. Five-year ACS data are available starting only in 2010, and because these estimates span a 5-year period, they are of limited utility in assessing the coverage of the NCVS sample for single-year estimates. For this reason, decennial Census data were used for comparisons with NCVS estimates.[1]

Table 2.1 contains definitions for each of the three subnational identifiers available on the NCVS PUF, including the levels and sources used to define each variable. Note that the levels of each variable are those used in the coverage evaluation and do not reflect the final collapsing of variables recommended for analysis, as discussed in ***Section 3.4***.

[1] In the future, if characteristics such as employment, education, and marital status would be useful in calibration models, ACS microdata could be considered for poststratification purposes.

Table 2.1. Generic area variable definitions

Generic area variable[1]	PUF variable	Generic area variable levels	Generic area variable sources
Census region	V2127B	1 = Northeast 2 = Midwest 3 = South 4 = West	Census region classification (defined at the state level)
Population size	V2126B	1 = Not in a place 2 = < 10,000 3 = 10,000–49,999 4 = 50,000–99,999 5 = 100,000–249,999 6 = 250,000–499,999 7 = 500,000–999,999 8 = 1,000,000 +	Census place size code ▪ 1990 Census population for the 1996–2005 NCVS ▪ 2000 Census population for the 2006–2012 NCVS
Urbanicity	V2129	1 = Central or principal city of a MSA/CBSA (urban) 2 = In MSA/CBSA, but not in the central or principal city (suburban) 3 = Not in an MSA/CBSA (rural)	CBSA/MSA status ▪ 1993 MSA and central city classifications for the 1996–2005 NCVS ▪ 2003 CBSA[2] and principal city classifications for the 2006–2012 NCVS

Notes: CBSA = Core Based Statistical Area; MSA = Metropolitan Statistical Area; NCVS = National Crime Victimization Survey; PUF = public-use file.

[1] Census region and population size were not included on the PUF until quarter 3 of 1995.

[2] Only metropolitan areas are classified as CBSAs; micropolitan areas are considered not in a CBSA.

Census region is classified at the State level and is thus fixed over time, but population size and urbanicity are redefined periodically on the basis of population growth in local areas (Census places and Metropolitan Statistical Areas [MSAs]/Core Based Statistical Areas [CBSAs]). The 1996–2005 NCVS samples are based on the 1990 Census and use 1993 MSA classifications, and the 2006–2012 samples are based on the 2000 Census and use 2003 CBSA classifications.[2] This means that the population size and urbanicity classifications for the NCVS sample are fixed for a 10-year period regardless of the growth in that area during that period or any updates made to MSA/CBSA classifications during that time.

[2] Years in the phase-in/phase-out periods contain some cases based on the prior years' design and others based on the new design, and therefore have variables that are defined with multiple Census definitions.

Whereas geographic variables are assigned to NCVS respondents on the basis of the Census used to develop the primary sampling units during sampling, NCVS weights are calibrated based on the most recent Census data available. Therefore, calculating comparable Census population estimates in generic areas requires a two-step process:

1. Map the generic area classifications from the decennial Census used to select the sample to the most recent decennial Census data available for the calculation of population estimates; and

2. Using intercensal population projections, adjust decennial Census estimates to account for population changes between the most recent decennial Census data and the year of estimation.

This process varies depending on the decennial Census used to select the sample, the most recent decennial Census data year, and the number of years between the most recent decennial Census data year and the year of estimation. The coverage of the NCVS sample and ultimately the estimates themselves need to be assessed for multiple time periods to fully understand how well estimates within generic areas perform, so the coverage evaluation included three NCVS years that vary on the basis of the above characteristics:

1. A year that is based on the 1990 decennial Census but occurs after the 2000 decennial Census (2002);

2. A year that is based on the 2000 decennial Census but occurs before the 2010 decennial Census (2008); and

3. A year that is based on the 2000 decennial Census but occurs after the 2010 decennial Census (2012).

Comparable Census estimates for these 3 years were calculated by applying the two-step process outlined above. First, generic area variables were mapped from the decennial Census used to select the NCVS sample to the most recent decennial Census population totals available. Each of the three generic area variables is assigned at a different geographic level, so each was mapped to the most recent decennial Census using a different geographic crosswalk. The

assignment level of each variable, the mapping years, and the crosswalks used to map geographic variables across decennial Census years are outlined in ***Tables 2.2a–2.2c*** for the 2002, 2008, and 2012 estimates, respectively.

Table 2.2a. 2002 National Crime Victimization Survey generic area classifications and mappings

Generic area variable	Assignment level	Mapping	Crosswalks used
Census region	State	1990–2000	N/A; Census region definitions did not change from 1990 to 2000
Population size	Census place	1990–2000	RTI-created 1990 to 2000 Census place crosswalk[1]
Urbanicity	MSAs: county or county subdivision Central cities: Census place	1990–2000	Publicly available 1990 to 2000 county crosswalk[2] RTI-created 1990 to 2000 county subdivision crosswalk[1] RTI-created 1990 to 2000 Census place crosswalk[1]

Note: MSA = Metropolitan Statistical Area.

[1] RTI developed a crosswalk from 1990 to 2000 places and county subdivisions by overlaying the Census place and county subdivision shape files from the 1990 Census onto the 2000 Census blocks and assigning the Census place and county subdivision associated with the centroid of each 2000 Census block.

[2] Available from https://www.census.gov/geo/reference/county-changes.html

Table 2.2b. 2008 National Crime Victimization Survey generic area classifications and mappings

Generic area variable	Assignment level	Mapping[1]	Crosswalks used
Census region	State	N/A	N/A
Population size	Census place	N/A	N/A
Urbanicity	CBSAs: county Principal cities: Census place	N/A	N/A

Note: CBSA = Core Based Statistical Area.

[1] For the 2008 NCVS, the decennial Census used during sampling is the most recent decennial Census data available, so no mapping is required.

Table 2.2c. 2012 National Crime Victimization Survey generic area classifications and mappings

Generic area variable	Assignment level	Mapping	Crosswalks used
Census region	State	2000–2010	N/A; Census region definitions did not change from 2000 to 2010
Population size	Census place	2000–2010	Publicly available 2000 to 2010 Census place crosswalk[1]
Urbanicity	CBSAs: county Principal cities: Census place	2000–2010	Publicly available 2000 to 2010 county crosswalk[2] Publicly available 2000 to 2010 Census place crosswalk[1]

Note: CBSA = Core Based Statistical Area.

[1] Available from https://www.census.gov/geo/maps-data/data/comp-place.html

[2] Available from https://www.census.gov/geo/reference/county-changes.html

After generic areas were mapped from the decennial Census that was used to select the sample to the most recent decennial Census for obtaining estimates (step 1), decennial Census population totals were adjusted to account for population changes between the most recent decennial Census data available and the year of estimation (step 2). Adjustment factors were calculated for each estimation year by taking the ratio of the intercensal annual projected population counts for households and persons at the Census region level[3] (*PROJHH* and *PROJPER*) to decennial estimated population counts of households and persons from the most recent Census (*HH* and *PER*). The household and person adjustment factors for Census region *i*, estimation year *j*, and decennial Census reference year *k*, are

$$HHADJ_{ijk} = \frac{PROJHH_{ij}}{HH_{ik}} \text{ and } PERADJ_{ijk} = \frac{PROJPER_{ij}}{PER_{ik}}$$

The decennial Census reference year is the most recent Census year associated with the estimation year (i.e., 2000 for the 2002 and 2008 estimation years and 2010 for the 2012 estimation year). These adjustment factors inflate the estimates from the decennial Census reference year up to the estimation year and were applied to household- and person-level demographic characteristics to obtain the gold-standard Census population totals in each generic area for comparison with the NCVS estimates.

[3] Available from https://www.census.gov/popest/data/historical/index.html

After the implementation of step 1 and step 2 as described above, Census estimates were calculated within generic areas that were comparable to the NCVS generic areas. Estimated totals and population distributions were calculated from the Census for the person- and household-level demographic characteristics listed below for 2002, 2008, and 2012. These characteristics were available in each decennial Census short form.

Person-level characteristics:

- Gender
- Age category
- Race/ethnicity
- Number of persons in the household
- Household tenure

Household-level characteristics:

- Age of householder
- Race/ethnicity of householder
- Number of persons in the household
- Household tenure

2.3 Coverage Evaluation

To assess how well the NCVS population totals in generic areas aligned with the true gold-standard Census population totals, NCVS population estimates were calculated in the 114 generic areas[4] for the same set of person- and household-level characteristics listed in ***Section 2.2***.

Within each of the 114 generic areas, NCVS population estimates were compared to Census gold-standard estimates. For each person- and household-level demographic characteristic analyzed, significance testing was conducted to determine whether the NCVS and

[4] The 114 generic areas comprised 56 region × population size × urbanicity areas, 32 region × population size areas, 12 region × urbanicity areas, and 14 urbanicity × population size areas. The number of generic areas is not always equal to the total number of combinations of the geographic variables because some variable combinations do not exist in the population (e.g., urban areas that are not in a Census place).

Census estimated totals were significantly different at the 95 percent confidence level. These significance tests assumed that the NCVS and Census estimates are independent and that the Census estimates have no variation. ***Tables 2.3a–2.3d*** summarize the number of statistically significant differences across each of the two- or three-variable generic area groups by demographic characteristic.

In general, the NCVS estimates align well with the Census estimates. However, there were a nonnegligible number of statistically significant differences across all four types of generic areas. The differences vary slightly for the three survey years and vary based on the demographic characteristic of interest (e.g., householder age tends to align well, whereas the number of persons in the household does not align as well). In addition, differences were not uniform across generic area types. Census region × urbanicity NCVS totals align better with Census estimates compared with the other three types of generic areas.

Because the differences between the NCVS estimated population demographic distributions and the Census estimates in the generic areas were statistically significant for a nonnegligible number of comparisons, further assessment was needed to determine whether estimates produced with the original NCVS weights would exhibit bias. Differences in population totals do not necessarily indicate that NCVS estimates calculated with the original NCVS weights will be biased. Slight differences in the population distributions do not necessarily equate to differences in resulting victimization totals and rates. Further analysis was needed to determine whether and how calibrating NCVS weights to Census control totals within generic areas would affect key estimates.

Table 2.3a. Comparison of Census and NCVS population estimates by demographic characteristics: region × population size × urbanicity

	2002		2008		2012	
	Num diff.[a]	Pct diff.	Num diff.[a]	Pct diff.	Num diff.[a]	Pct diff.
Person-level estimates						
All persons 12+	9	16.1%	8	14.3%	11	19.6%
Gender						
Male	8	14.3%	7	12.5%	10	17.9%
Female	7	12.5%	8	14.3%	7	12.5%
Age						
12–15	2	3.6%	7	12.5%	7	12.5%
16–19	9	16.1%	11	19.6%	7	12.5%
20–24	6	10.7%	8	14.3%	5	8.9%
25–34	13	23.2%	8	14.3%	8	14.3%
35–49	7	12.5%	9	16.1%	10	17.9%
50–64	0	0.0%	13	23.2%	7	12.5%
65 or older	7	12.5%	12	21.4%	4	7.1%
Race						
White[b]	7	12.5%	5	8.9%	12	21.4%
Black[b]	11	19.6%	10	17.9%	10	17.9%
Other[c]	52	92.9%	34	60.7%	38	67.9%
Ethnicity						
Hispanic	35	62.5%	21	37.5%	32	57.1%
Non-Hispanic	8	14.3%	5	8.9%	14	25.0%
Number of persons in the household						
1	56	100.0%	54	96.4%	18	32.1%
2	24	42.9%	24	42.9%	10	17.9%
3+	36	64.3%	32	57.1%	13	23.2%
Household tenure						
Own	3	5.4%	2	3.6%	7	12.5%
Rent/no cash rent	13	23.2%	13	23.2%	7	12.5%

(continued)

Table 2.3a. Comparison of Census and NCVS population estimates by demographic characteristics: region × population size × urbanicity (continued)

	2002		2008		2012	
	Num diff.[a]	Pct diff.	Num diff.[a]	Pct diff.	Num diff.[a]	Pct diff.
Household-level estimates						
All households	4	7.1%	5	8.9%	9	16.1%
Age of householder						
Younger than 25	5	8.9%	8	14.3%	14	25.0%
25–34	2	3.6%	3	5.4%	9	16.1%
35–44	4	7.1%	12	21.4%	7	12.5%
45–54	0	0.0%	8	14.3%	5	8.9%
55–64	2	3.6%	16	28.6%	6	10.7%
65 or older	6	10.7%	11	19.6%	6	10.7%
Race/ethnicity of householder						
White[b]	4	7.1%	7	12.5%	5	8.9%
Black[b]	4	7.1%	7	12.5%	10	17.9%
Hispanic	11	19.6%	14	25.0%	12	21.4%
Other/more than one race[b,c]	21	37.5%	6	10.7%	9	16.1%
Number of persons in the household						
1	0	0.0%	9	16.1%	15	26.8%
2	1	1.8%	5	8.9%	12	21.4%
3+	9	16.1%	12	21.4%	10	17.9%
Household tenure						
Own	2	3.6%	3	5.4%	9	16.1%
Rent/no cash rent	5	8.9%	8	14.3%	8	14.3%

Note: NCVS = National Crime Victimization Survey.

[a] Number of areas (of 56 total) where NCVS and Census estimated totals are statistically significant at the 95% confidence level.

[b] Excludes those of Hispanic origin.

[c] Includes Asian, Native Hawaiian, Other Pacific Islander, American Indian, Alaska Native, and persons with two or more races.

Table 2.3b. Comparison of Census and NCVS population estimates by demographic characteristics: region × population size

	2002		2008		2012	
	Num diff.[a]	Pct diff.	Num diff.[a]	Pct diff.	Num diff.[a]	Pct diff.
Person-level estimates						
All persons 12+		25.0%	6	18.8%	7	21.9%
Gender						
Male	7	21.9%	4	12.5%	6	18.8%
Female	6	18.8%	6	18.8%	5	15.6%
Age						
12–15	1	3.1%	2	6.3%	3	9.4%
16–19	4	12.5%	4	12.5%	3	9.4%
20–24	2	6.3%	3	9.4%	3	9.4%
25–34	10	31.3%	6	18.8%	4	12.5%
35–49	6	18.8%	5	15.6%	8	25.0%
50–64	1	3.1%	11	34.4%	4	12.5%
65 or older	7	21.9%	9	28.1%	3	9.4%
Race						
White[b]	5	15.6%	5	15.6%	11	34.4%
Black[b]	3	9.4%	7	21.9%	3	9.4%
Other[b,c]	30	93.8%	23	71.9%	26	81.3%
Ethnicity						
Hispanic	24	75.0%	14	43.8%	21	65.6%
Non-Hispanic	5	15.6%	4	12.5%	10	31.3%
Number of persons in the household						
1	32	100.0%	32	100.0%	19	59.4%
2	14	43.8%	18	56.3%	8	25.0%
3+	29	90.6%	29	90.6%	12	37.5%
Household tenure						
Own	3	9.4%	1	3.1%	4	12.5%
Rent/no cash rent	10	31.3%	10	31.3%	7	21.9%

(continued)

Table 2.3b. Comparison of Census and NCVS population estimates by demographic characteristics: region × population size (continued)

	2002		2008		2012	
	Num diff.[a]	Pct diff.	Num diff.[a]	Pct diff.	Num diff.[a]	Pct diff.
Household-level estimates						
All households		15.6%	5	15.6%	9	28.1%
Age of householder						
Younger than 25	2	6.3%	4	12.5%	12	37.5%
25–34	2	6.3%	3	9.4%	6	18.8%
35–44	3	9.4%	11	34.4%	3	9.4%
45–54	1	3.1%	7	21.9%	4	12.5%
55–64	2	6.3%	12	37.5%	3	9.4%
65 or older	6	18.8%	8	25.0%	7	21.9%
Race/ethnicity of householder						
White[b]	3	9.4%	5	15.6%	4	12.5%
Black[b]	2	6.3%	5	15.6%	6	18.8%
Hispanic	6	18.8%	11	34.4%	6	18.8%
Other/more than one race[b,c]	14	43.8%	3	9.4%	2	6.3%
Number of persons in the household						
1	0	0.0%	9	28.1%	14	43.8%
2	2	6.3%	6	18.8%	8	25.0%
3+	7	21.9%	11	34.4%	7	21.9%
Household tenure						
Own	2	6.3%	5	15.6%	7	21.9%
Rent/no cash rent	3	9.4%	6	18.8%	7	21.9%

Note: NCVS = National Crime Victimization Survey.

[a] Number of areas (of 56 total) where NCVS and Census estimated totals are statistically significant at the 95% confidence level.

[b] Excludes those of Hispanic origin.

[c] Includes Asian, Native Hawaiian, Other Pacific Islander, American Indian, Alaska Native, and persons with two or more races.

Table 2.3c. Comparison of Census and NCVS population estimates by demographic characteristics: region × urbanicity

	2002		2008		2012	
	Num diff.[a]	Pct diff.	Num diff.[a]	Pct diff.	Num diff.[a]	Pct diff.
Person-level estimates						
All persons 12+		0.0%	1	8.3%	0	0.0%
Gender						
Male	0	0.0%	2	16.7%	0	0.0%
Female	0	0.0%	1	8.3%	0	0.0%
Age						
12–15	0	0.0%	1	8.3%	1	8.3%
16–19	1	8.3%	1	8.3%	1	8.3%
20–24	1	8.3%	1	8.3%	0	0.0%
25–34	3	25.0%	1	8.3%	0	0.0%
35–49	0	0.0%	4	33.3%	1	8.3%
50–64	0	0.0%	7	58.3%	0	0.0%
65 or older	3	25.0%	1	8.3%	0	0.0%
Race						
White[b]	3	25.0%	1	8.3%	3	25.0%
Black[b]	2	16.7%	3	25.0%	1	8.3%
Other[b,c]	11	91.7%	7	58.3%	10	83.3%
Ethnicity						
Hispanic	11	91.7%	5	41.7%	11	91.7%
Non-Hispanic	1	8.3%	0	0.0%	3	25.0%
Number of persons in the household						
1	12	100.0%	12	100.0%	8	66.7%
2	4	33.3%	6	50.0%	5	41.7%
3+	8	66.7%	10	83.3%	6	50.0%
Household tenure						
Own	0	0.0%	1	8.3%	0	0.0%
Rent/no cash rent	3	25.0%	2	16.7%	1	8.3%

(continued)

Table 2.3c. Comparison of Census and NCVS population estimates by demographic characteristics: region × urbanicity (continued)

	2002		2008		2012	
	Num diff.[a]	Pct diff.	Num diff.[a]	Pct diff.	Num diff.[a]	Pct diff.
Household-level estimates						
All households		0.0%	0	0.0%	2	16.7%
Age of householder						
Younger than 25	7	58.3%	5	41.7%	8	66.7%
25–34	0	0.0%	0	0.0%	5	41.7%
35–44	0	0.0%	7	58.3%	0	0.0%
45–54	0	0.0%	2	16.7%	0	0.0%
55–64	1	8.3%	7	58.3%	0	0.0%
65 or older	0	0.0%	0	0.0%	0	0.0%
Race/ethnicity of householder						
White[b]	0	0.0%	0	0.0%	0	0.0%
Black[b]	3	25.0%	4	33.3%	7	58.3%
Hispanic	4	33.3%	7	58.3%	5	41.7%
Other/more than one race[b,c]	11	91.7%	3	25.0%	1	8.3%
Number of persons in the household						
1	0	0.0%	6	50.0%	8	66.7%
2	1	8.3%	4	33.3%	6	50.0%
3+	0	0.0%	2	16.7%	0	0.0%
Household tenure						
Own	0	0.0%	2	16.7%	1	8.3%
Rent/no cash rent	1	8.3%	0	0.0%	1	8.3%

Note: NCVS = National Crime Victimization Survey.

[a] Number of areas (of 56 total) where NCVS and Census estimated totals are statistically significant at the 95% confidence level.

[b] Excludes those of Hispanic origin.

[c] Includes Asian, Native Hawaiian, Other Pacific Islander, American Indian, Alaska Native, and persons with two or more races.

Table 2.3d. Comparison of Census and NCVS population estimates by demographic characteristics: population size × urbanicity

	2002		2008		2012	
	Num diff.[a]	Pct diff.	Num diff.[a]	Pct diff.	Num diff.[a]	Pct diff.
Person-level estimates						
All persons 12+		28.6%	2	14.3%	1	7.1%
Gender						
Male	3	21.4%	2	14.3%	1	7.1%
Female	4	28.6%	2	14.3%	1	7.1%
Age						
12–15	0	0.0%	2	14.3%	1	7.1%
16–19	1	7.1%	2	14.3%	1	7.1%
20–24	0	0.0%	0	0.0%	1	7.1%
25–34	5	35.7%	3	21.4%	3	21.4%
35–49	3	21.4%	4	28.6%	3	21.4%
50–64	0	0.0%	5	35.7%	1	7.1%
65 or older	4	28.6%	5	35.7%	1	7.1%
Race						
White[b]	3	21.4%	2	14.3%	5	35.7%
Black[b]	0	0.0%	4	28.6%	1	7.1%
Other[b,c]	14	100.0%	13	92.9%	14	100.0%
Ethnicity						
Hispanic	10	71.4%	9	64.3%	12	85.7%
Non-Hispanic	3	21.4%	2	14.3%	5	35.7%
Number of persons in the household						
1	14	100.0%	14	100.0%	10	71.4%
2	8	57.1%	7	50.0%	4	28.6%
3+	12	85.7%	13	92.9%	5	35.7%
Household tenure						
Own	1	7.1%	1	7.1%	1	7.1%
Rent/no cash rent	5	35.7%	3	21.4%	1	7.1%

(continued)

Table 2.3d. Comparison of Census and NCVS population estimates by demographic characteristics: population size × urbanicity (continued)

	2002		2008		2012	
	Num diff.[a]	Pct diff.	Num diff.[a]	Pct diff.	Num diff.[a]	Pct diff.
Household-level estimates						
All households		0.0%	1	7.1%	4	28.6%
Age of householder						
Younger than 25	6	42.9%	5	35.7%	8	57.1%
25–34	2	14.3%	0	0.0%	4	28.6%
35–44	1	7.1%	9	64.3%	1	7.1%
45–54	0	0.0%	2	14.3%	1	7.1%
55–64	0	0.0%	7	50.0%	4	28.6%
65 or older	1	7.1%	2	14.3%	2	14.3%
Race/ethnicity of householder						
White[b]	1	7.1%	2	14.3%	3	21.4%
Black[b]	2	14.3%	2	14.3%	5	35.7%
Hispanic	3	21.4%	7	50.0%	6	42.9%
Other/more than one race[b,c]	12	85.7%	1	7.1%	0	0.0%
Number of persons in the household						
1	2	14.3%	4	28.6%	7	50.0%
2	2	14.3%	3	21.4%	5	35.7%
3+	5	35.7%	4	28.6%	3	21.4%
Household tenure						
Own	1	7.1%	2	14.3%	4	28.6%
Rent/no cash rent	1	7.1%	0	0.0%	3	21.4%

Note: NCVS = National Crime Victimization Survey.

[a] Number of areas (of 56 total) where NCVS and Census estimated totals are statistically significant at the 95% confidence level.

[b] Excludes those of Hispanic origin.

[c] Includes Asian, Native Hawaiian, Other Pacific Islander, American Indian, Alaska Native, and persons with two or more races.

SECTION 3. WEIGHTING EVALUATION

3.1 Overview of Weighting Evaluation

On the basis of the results of the coverage analysis, NCVS weights were calibrated to Census control totals at the generic area level, and key victimization rates and totals were calculated both with the original NCVS weights and the calibrated weights. Comparisons were made between the two sets of weights based on the estimates themselves and the precision of estimates. Calibrating the NCVS weights to Census control totals should minimize potential bias in estimates at the generic-area level. However, because calibration models often increase the variation in analysis weights, calibrated estimates are likely to be less precise than estimates based on the original NCVS weights. The weight calibration approach and the results of these comparisons are discussed in the sections below. In addition, this section compares the precision of single- and pooled-year estimates, evaluates the effects of collapsing within generic area levels on resulting precision, and makes recommendations for the most appropriate weighting approach for generic area estimation.

3.2 Calibration of NCVS Weights in Generic Areas

For each of the four generic area types, weight calibration models were fit for the household- and person-level NCVS data files for each analysis year on the basis of the set of characteristics that are commonly defined in the NCVS and the Census (those used in the coverage comparison tables). Because the gold-standard Census estimates were designed to accurately represent each generic area, these population counts can be used as control totals in the poststratification models. ***Table 3.1*** lists the characteristics that were used for person-level models, and ***Table 3.2*** lists the characteristics that were used for household-level models. The set of characteristics used in the person-level models includes both person- and household-level characteristics. The set of characteristics for the household-level models includes characteristics about the householder (i.e., the reference person in the NCVS) and the household.

Table 3.1. Characteristics for person-level poststratification model by number of levels

Characteristic	Number of levels
Gender of respondent	2
Age category of respondent	7
Race of respondent	3
Ethnicity of respondent	2
Number of persons in the household	3
Household tenure	2

Table 3.2. Characteristics for household-level poststratification model by number of levels

Characteristic	Number of levels
Age of householder	6
Race/ethnicity of householder	4
Household tenure	2
Number of persons in the household	3

To ensure that control totals were maintained for each generic area within each type (e.g., totals for Northeast/Urban area were maintained within the Census region × urbanicity model), calibration models included a variable specifying each generic area and the interactions between this variable and all demographic variables.

Calibration models were fit using generalized exponential modeling (GEM; available in SUDAAN's "WTADJUST" procedure) for single-year estimates (2002, 2008, and 2012) as well as 3-year pooled estimates (2000–2002, 2006–2008, and 2010–2012). For the 2002 and 2000–2002 models, the race of the respondent was excluded from the person-level models because the NCVS and Census population estimates were not defined comparably.[5] However, no further collapsing or removal of variables was required for the calibration models to converge.

[5] The 2000 Census classified respondents into race categories that included a "multiple races" option. The NCVS instrument for 2000–2002 did not allow respondents to select multiple races. For this reason, the Census and NCVS distributions by race are not comparable.

3.3 Comparison of Recalibrated Weights to Original Weights

After the original NCVS weights were calibrated to Census control totals for each generic area and analysis year, key person- and household-level victimization rates and totals (listed below) were calculated overall and for crimes reported to the police:

- All crime
- Violent crime
- Serious violent crime
- Rape/sexual assault
- Robbery
- Assault
- Aggravated assault
- Simple assault
- Personal theft
- Property crime
- Household burglary
- Motor vehicle theft
- Theft
- Violent crimes involving a weapon
- Violent crimes involving a firearm
- Violent crimes committed by a stranger
- Violent crimes committed by an intimate
- Violent crimes committed by other relative
- Violent crimes committed by other known offender
- Violent crimes occurring during the day
- Violent crimes occurring at night

Estimates were computed two ways: (1) on the basis of the original NCVS weights (uncalibrated estimates) and (2) on the basis of the generic-area specific adjusted weights (calibrated estimates). For each crime type, the estimates and relative standard errors (RSEs) were calculated using both sets of weights. The calibrated and uncalibrated victimization rates for all generic areas are compared in ***Figures 3.1*** and ***3.2*** for 1- and 3-year estimates, respectively. The figures on the left contain all rates computed (11,868 comparisons), whereas the figures on the right are limited to rates that were not flagged as unreliable on the basis of the sample size and RSE[6] (4,109 and 6,607 comparisons for 1- and 3-year estimates, respectively). For both 1- and 3-year estimates, calibrated and uncalibrated rates align nicely along the 45-degree line of equality, particularly when the estimates compared are limited to those with

[6] Estimates based on 10 or fewer victimizations or with RSEs greater than 30% are flagged as unreliable.

Figure 3.1. One-year victimization rates: Calibrated compared with uncalibrated estimates

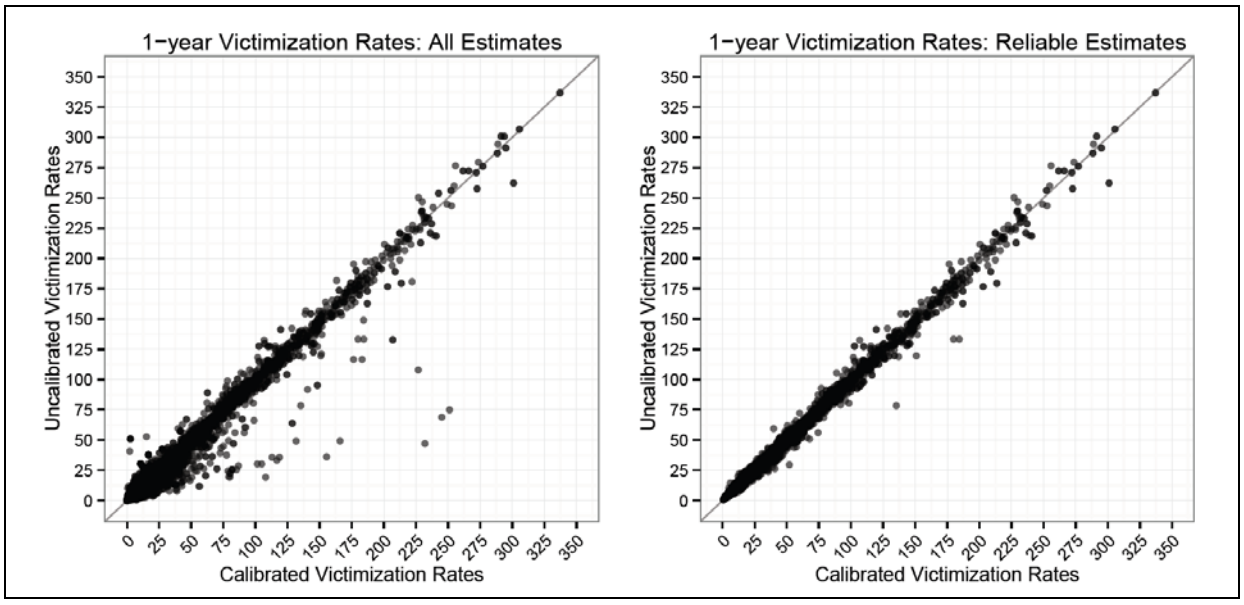
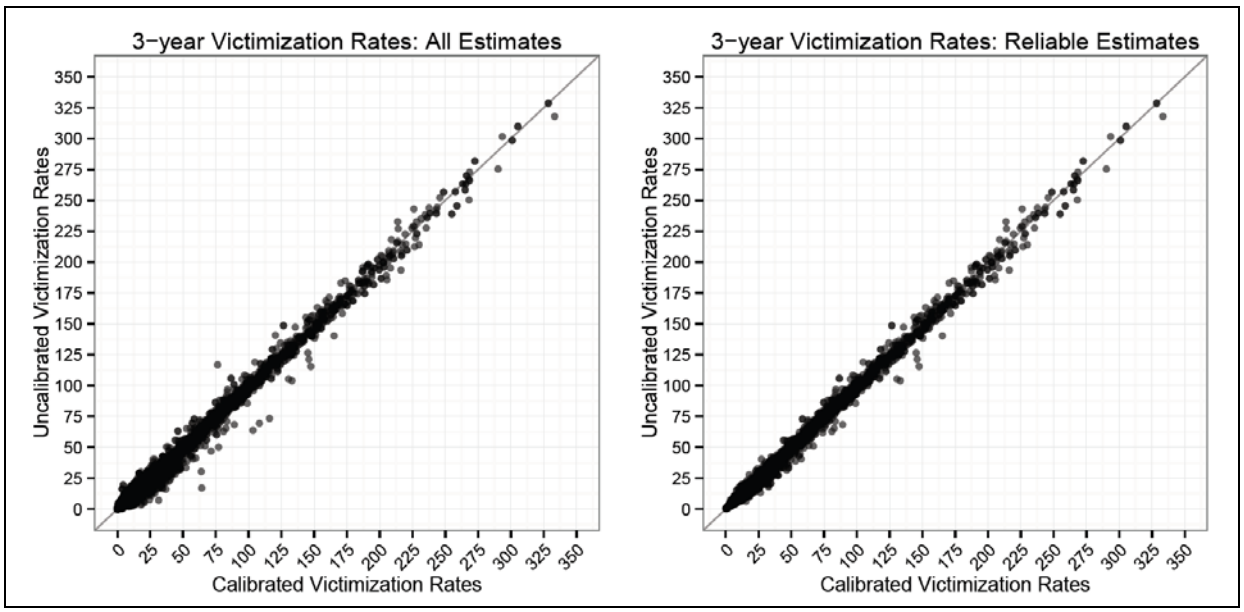


Figure 3.2. Three-year victimization rates: Calibrated compared with uncalibrated estimates



reasonable precision (i.e., reliable estimates). This provides evidence that weight calibration to Census control totals within generic areas did not lead to substantive differences in key NCVS victimization rates.

Similarly, the calibrated and uncalibrated victimization totals for all generic areas are compared in ***Figures 3.3*** and ***3.4*** for 1- and 3-year estimates, respectively. The figures on the left contain all totals (11,868 comparisons), whereas the figures on the right are limited to totals that were not flagged as unreliable (3,665 and 5,866 comparisons for 1- and 3-year estimates, respectively). As with the victimization rates, for both 1- and 3-year estimates, calibrated and uncalibrated totals align nicely along the 45-degree line of equality, providing further evidence that weight calibration to Census control totals within generic areas did not lead to substantive differences in key NCVS estimates.

Figure 3.3. One-year victimization totals: Calibrated compared with uncalibrated estimates

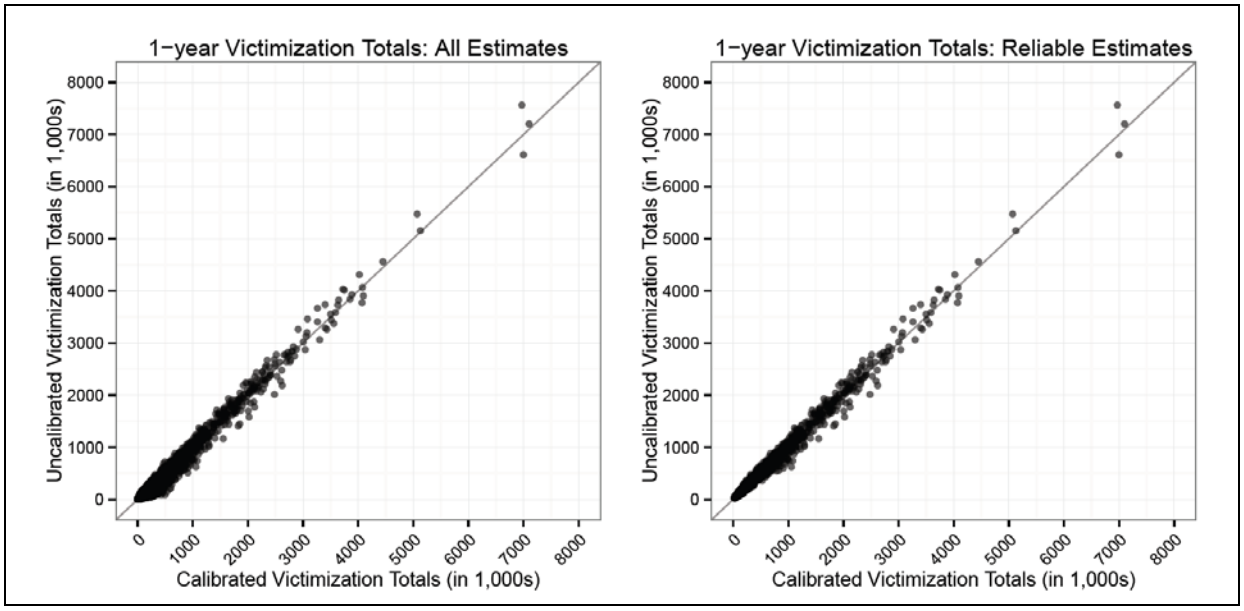
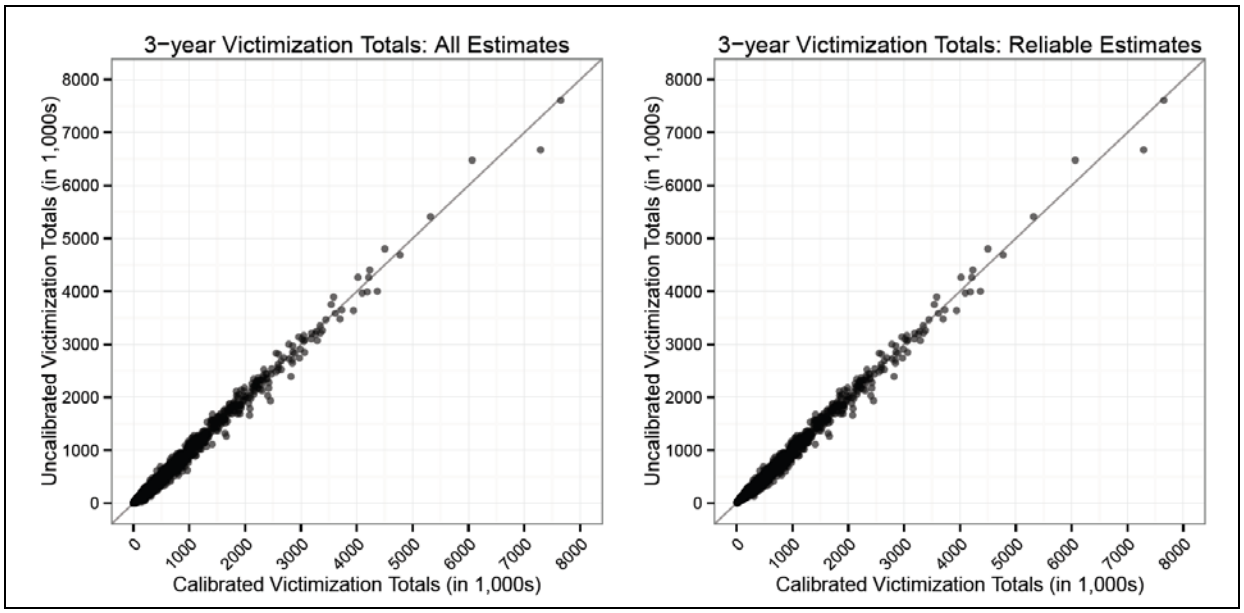


Figure 3.4. Three-year victimization totals: Calibrated compared with uncalibrated estimates



The impact of calibrating the NCVS weights on the precision of key victimization rates and totals was also assessed. As previously noted, calibrating estimates often adds variation to analysis weights, which makes resulting estimates less precise. The precision of calibrated and uncalibrated estimates for all generic areas is compared in ***Figures 3.5*** and ***3.6*** for victimization rates and totals, respectively. These figures plot the RSEs for calibrated estimates against the RSEs for the corresponding uncalibrated estimates for 1- and 3-year periods, but they are limited to estimates for which both the calibrated and uncalibrated estimates had RSEs of less than 50 percent (8,076 and 9,707 comparisons for 1- and 3-year rates, respectively; 7,786 and 9,397 comparisons for 1- and 3-year totals, respectively), as estimates with RSEs above 50 percent are not reliable and do not provide useful comparisons.

For both 1- and 3-year estimates and for both victimization rates and totals, the majority of comparisons fall below the 45-degree line of equality, indicating that calibrated estimates have higher RSEs than uncalibrated estimates. For victimization rates, 55 percent of 1-year comparisons and 65 percent of 3-year comparisons fall below the 45-degree line, whereas for victimization totals, 54 percent of 1-year comparisons and 63 percent of 3-year comparisons fall below the 45-degree line. Thus, the calibration process resulted in a loss of precision for key NCVS estimates while not substantively affecting the estimates themselves.

Figure 3.5. Relative standard errors of calibrated compared with uncalibrated estimates: 1- and 3-year victimization rates

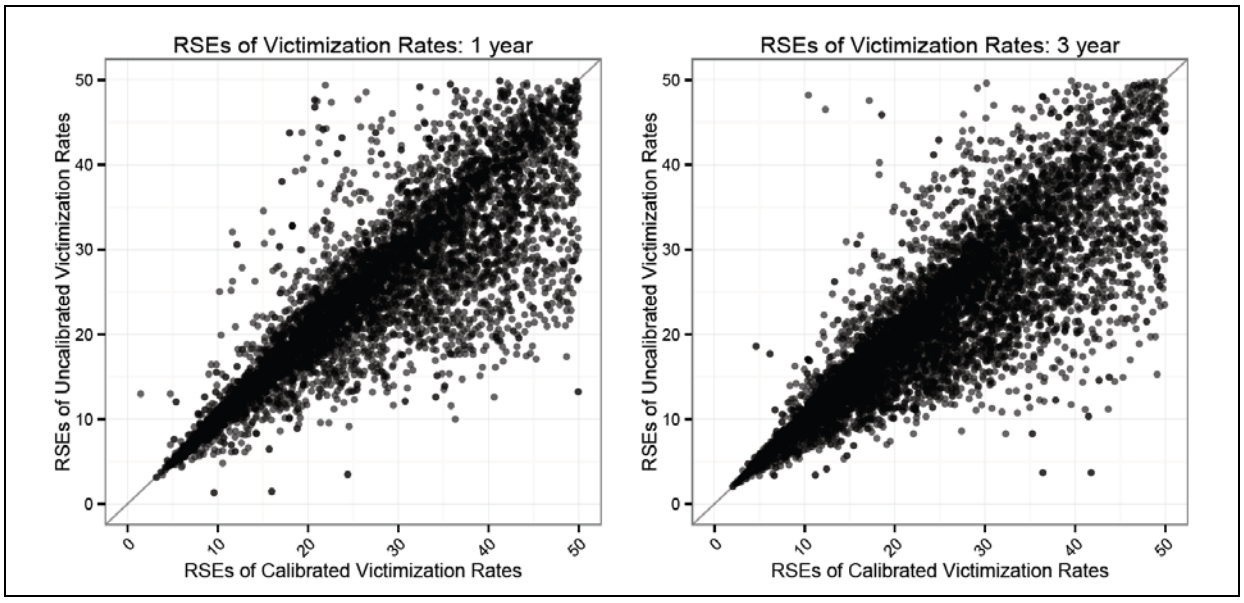
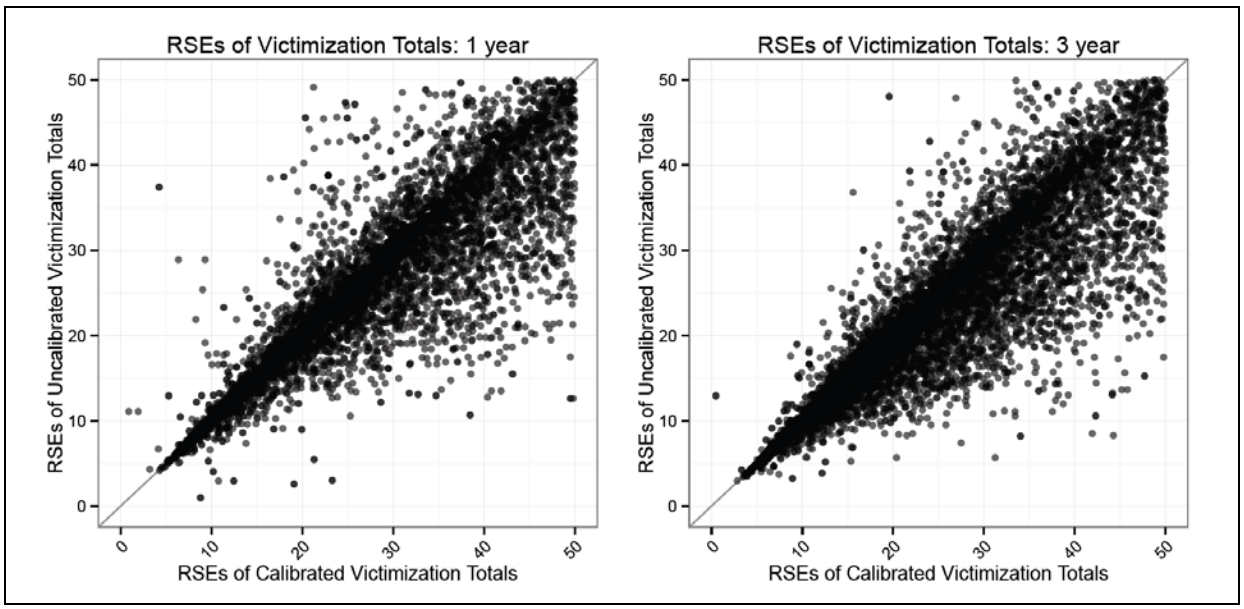


Figure 3.6. Relative standard errors of calibrated compared with uncalibrated estimates: 1- and 3-year victimization totals



3.4 Determining Appropriate Collapsing and Pooling of Data for Generic Area Estimation

In addition to assessing the best analytic weights to use within generic areas, the RTI team evaluated the best approach for defining generic areas and pooling data across years to achieve precise but timely estimates within generic areas. One-year estimates provide the timeliest data within generic areas, but they are less precise than 3- or 5-year estimates because of smaller annual sample sizes. Similarly, collapsing levels within generic areas provides added precision, particularly for domains with small sample sizes. During the weighting evaluation, several of the generic areas had low levels of precision for key estimates, especially those defined within the Census region × population size × urbanicity generic area type. To improve precision levels across all generic areas such that key estimates would have reasonable precision, both variable collapsing and pooling of data across years were explored.

Because the population size variable has the most levels and can be collapsed into categories that are reasonable from a substantive perspective (i.e., it is not reasonable to collapse across Census regions or urbanicity categories), three potential collapsing options for population size were evaluated. The first option collapses population size into 6 levels that are consistent with the NCVS Victimization Analysis Tool (NVAT) and leads to a total of 40 Census region × population size × urbanicity categories. The second option further collapses the middle levels of population size, leading to 28 total Census region × population size × urbanicity categories. The third option instead collapses the highest levels of population size, leading to 32 total Census region × population size × urbanicity categories. These options were selected to ameliorate precision concerns for the levels of population size that had the lowest levels of precision in the weighting evaluation. The three collapsing options considered are presented in ***Table 3.3***.

Table 3.3. Population size collapsing options

Option 1	Option 2	Option 3
Not in a place	Not in a place	Not in a place
<100,000	<100,000–249,999	<100,000
100,000–249,999	250,000–999,999	100,000–499,999
250,000–499,999	1,000,000+	500,000+
500,000–999,999		
1,000,000+		

In addition to considering the three collapsed versions of population size, the RTI team also evaluated the effects of pooling 3 and 5 years of data to calculate victimization rates within generic areas. The percentage of 1-, 3-, and 5-year estimates that would be flagged as unreliable on the basis of sample sizes and percent RSEs was calculated for each of the three collapsing options. ***Tables 3.4a*** and ***3.4b*** present the percentage of estimates that would be flagged as unreliable on the basis of collapsing option 1 for all crimes and for crimes reported to police, respectively. ***Tables 3.5a*** and ***3.5b*** present similar estimates for collapsing option 2, and ***Tables 3.6a*** and ***3.6b*** present estimates for collapsing option 3. All tables include 2002, 2008, and 2012 estimates for 1-year periods; 2000–2002, 2006–2008, and 2010–2012 estimates for 3-year periods; and 1998–2002, 2004–2008, and 2008–2012 for 5-year periods.

Both four-level versions of population size (collapsing options 2 and 3) led to significantly fewer flagged estimates than the six-level NVAT version (collapsing option 1). Collapsing option 3 led to more reliable estimates for the Census region × population size generic areas, and collapsing option 2 led to more reliable estimates for the Census region × population size × urbanicity generic areas (most likely because it leads to fewer levels of this variable). Because the Census region × population size × urbanicity generic areas were the most problematic in terms of precision, it is recommended that collapsing option 2 be implemented because it provides the most precision gains for the most problematic generic area type.

The appropriate number of years of data to pool is dependent on the key estimate and the generic area types included in the analysis. In general, 3-year pooled estimates have reasonable precision for most types of crime for the two-variable generic area types (Census region × population size, Census region × urbanicity, and population size × urbanicity). Pooling 5 years of data is required to obtain reasonable precision for many crime types for the Census region × population size × urbanicity generic areas. However, even after collapsing the population size variable and pooling 5 years of data, some of the rarest crime types (e.g., rape and sexual assault, personal theft, violent crimes committed by other relatives) do not have adequate precision for the majority of generic areas and would require further pooling of years or further collapsing of generic area levels to produce estimates with reasonable quality.

Table 3.4a. Percentage of victimization rates flagged as unreliable by type of crime: Option 1—Crimes reported and not reported to police

Type of crime	1-year				3-year				5-year			
	Reg. × pop. size	Reg. × urban-icity	Urban-icity × pop. size	Reg. × pop. size × urban-icity	Reg. × pop. size	Reg. × urban-icity	Urban-icity × pop. size	Reg. × pop. size × urban-icity	Reg. × Pop. Size	Reg. × urban-icity	Urban-icity × pop. size	Reg. × pop. size × urban-icity
All crime	1.4	0.0	0.0	7.5	0.0	0.0	0.0	4.2	0.0	0.0	0.0	1.7
Violent crime	27.8	8.3	3.3	43.3	11.1	5.6	0.0	17.5	4.2	2.8	0.0	11.7
Serious violent crime	59.7	30.6	23.3	78.3	26.4	5.6	0.0	39.2	11.1	2.8	0.0	21.7
Rape/sexual assault	100.0	97.2	100.0	100.0	94.4	63.9	56.7	96.7	83.3	47.2	36.7	89.2
Robbery	86.1	69.4	43.3	97.5	45.8	30.6	20.0	67.5	33.3	22.2	10.0	54.2
Assault	34.7	5.6	3.3	49.2	15.3	5.6	0.0	23.3	8.3	2.8	0.0	15.8
Aggravated	86.1	55.6	40.0	94.2	41.7	8.3	0.0	55.0	22.2	11.1	3.3	33.3
Simple	44.4	11.1	3.3	57.5	19.4	8.3	0.0	27.5	11.1	2.8	0.0	17.5
Personal theft	100.0	97.2	90.0	100.0	83.3	44.4	56.7	93.3	66.7	36.1	36.7	82.5
Property crime	2.8	0.0	0.0	7.5	0.0	0.0	0.0	2.5	0.0	0.0	0.0	0.8
Household burglary	20.8	2.8	0.0	36.7	8.3	0.0	0.0	17.5	4.2	0.0	0.0	7.5
Motor vehicle theft	65.3	36.1	36.7	82.5	19.4	11.1	0.0	45.0	6.9	11.1	0.0	23.3
Theft	4.2	0.0	0.0	8.3	1.4	0.0	0.0	4.2	0.0	0.0	0.0	1.7
Violent crimes involving a weapon	73.6	52.8	30.0	89.2	33.3	16.7	0.0	48.3	19.4	13.9	0.0	30.8
Violent crimes involving a firearm	93.1	75.0	66.7	99.2	68.1	38.9	30.0	85.0	41.7	25.0	3.3	65.0
Violent crimes committed by a stranger	52.8	27.8	0.0	68.3	22.2	8.3	0.0	35.0	9.7	2.8	0.0	20.0
Violent crimes committed by an intimate	91.7	88.9	70.0	99.2	75.0	30.6	26.7	85.0	51.4	8.3	3.3	60.8
Violent crimes committed by other relative	97.2	94.4	96.7	100.0	86.1	55.6	56.7	91.7	70.8	36.1	33.3	83.3
Violent crimes committed by other known offender	69.4	27.8	20.0	84.2	34.7	8.3	3.3	43.3	25.0	5.6	0.0	28.3
Violent crimes occurring during the day	48.6	19.4	3.3	64.2	22.2	8.3	0.0	34.2	15.3	8.3	0.0	25.0
Violent crimes occurring at night	45.8	19.4	10.0	59.2	20.8	8.3	0.0	30.8	11.1	2.8	0.0	23.3
HH crimes occurring during the day	8.3	5.6	0.0	20.0	1.4	0.0	0.0	4.2	0.0	0.0	0.0	3.3
HH crimes occurring at night	9.7	2.8	0.0	17.5	4.2	0.0	0.0	6.7	1.4	0.0	0.0	4.2

Notes: Estimates based on 10 or fewer sample cases, or with relative standard errors greater than 30%, are flagged as unreliable. HH = household.

Table 3.4b. Percentage of victimization rates flagged as unreliable by type of crime: Option 1—crimes reported to police

Type of crime	1-year				3-year				5-year			
	Reg. × pop. size	Reg. × urban-icity	Urban-icity × pop. size	Reg. × pop. size × urban-icity	Reg. × pop. size	Reg. × urban-icity	Urban-icity × pop. size	Reg. × pop. size × urban-icity	Reg. × Pop. Size	Reg. × urban-icity	Urban-icity × pop. size	Reg. × pop. size × urban-icity
All crime	9.7	2.8	0.0	17.5	0.0	0.0	0.0	5.0	0.0	0.0	0.0	0.8
Violent crime	45.8	16.7	20.0	56.7	20.8	8.3	0.0	33.3	9.7	2.8	0.0	21.7
Serious violent crime	79.2	52.8	23.3	90.0	27.8	8.3	3.3	42.5	18.1	8.3	0.0	30.8
Rape/sexual assault	100.0	100.0	100.0	100.0	98.6	83.3	96.7	100.0	93.1	66.7	86.7	97.5
Robbery	97.2	77.8	73.3	99.2	72.2	50.0	30.0	85.8	41.7	33.3	20.0	64.2
Assault	55.6	22.2	16.7	65.0	30.6	8.3	0.0	40.0	11.1	5.6	0.0	23.3
Aggravated	91.7	72.2	70.0	96.7	52.8	16.7	16.7	68.3	30.6	8.3	3.3	47.5
Simple	73.6	38.9	20.0	86.7	38.9	5.6	0.0	49.2	18.1	5.6	0.0	28.3
Personal theft	100.0	100.0	100.0	100.0	98.6	88.9	80.0	99.2	87.5	55.6	63.3	96.7
Property crime	11.1	0.0	0.0	19.2	2.8	0.0	0.0	5.8	0.0	0.0	0.0	0.8
Household burglary	36.1	16.7	3.3	55.8	9.7	0.0	0.0	16.7	8.3	0.0	0.0	12.5
Motor vehicle theft	73.6	44.4	36.7	85.0	23.6	16.7	0.0	48.3	5.6	13.9	0.0	27.5
Theft	15.3	2.8	0.0	25.8	6.9	0.0	0.0	10.0	2.8	0.0	0.0	5.8
Violent crimes involving a weapon	87.5	61.1	50.0	95.8	41.7	11.1	13.3	61.7	27.8	11.1	0.0	42.5
Violent crimes involving a firearm	100.0	86.1	90.0	99.2	79.2	52.8	33.3	90.8	51.4	27.8	13.3	78.3
Violent crimes committed by a stranger	76.4	44.4	23.3	90.0	41.7	13.9	6.7	52.5	15.3	8.3	0.0	30.8
Violent crimes committed by an intimate	95.8	88.9	93.3	100.0	80.6	52.8	43.3	94.2	68.1	13.9	26.7	81.7
Violent crimes committed by other relative	100.0	97.2	96.7	100.0	86.1	66.7	70.0	96.7	77.8	44.4	46.7	87.5
Violent crimes committed by other known offender	87.5	72.2	56.7	95.8	55.6	25.0	16.7	65.8	41.7	22.2	6.7	55.8
Violent crimes occurring during the day	75.0	44.4	40.0	90.8	38.9	16.7	3.3	50.8	23.6	8.3	0.0	33.3
Violent crimes occurring at night	79.2	38.9	20.0	88.3	34.7	16.7	0.0	43.3	22.2	8.3	0.0	31.7
HH crimes occurring during the day	25.0	13.9	3.3	39.2	9.7	0.0	0.0	15.0	6.9	0.0	0.0	10.0
HH crimes occurring at night	18.1	11.1	0.0	32.5	5.6	0.0	0.0	11.7	2.8	0.0	0.0	7.5

Notes: Estimates based on 10 or fewer sample cases, or with relative standard errors greater than 30%, are flagged as unreliable. HH = household.

Table 3.5a. Percentage of victimization rates flagged as unreliable by type of crime: Option 2—crimes reported and not reported to police

Type of crime	1-year				3-year				5-year			
	Reg. × pop. size	Reg. × urban-icity	Urban-icity × pop. size	Reg. × pop. size × urban-icity	Reg. × pop. size	Reg. × urban-icity	Urban-icity × pop. size	Reg. × pop. size × urban-icity	Reg. × Pop. Size	Reg. × urban-icity	Urban-icity × pop. size	Reg. × pop. size × urban-icity
All crime	0.0	0.0	0.0	3.6	0.0	0.0	0.0	2.4	0.0	0.0	0.0	0.0
Violent crime	18.8	8.3	0.0	33.3	4.2	5.6	0.0	11.9	0.0	2.8	0.0	6.0
Serious violent crime	50.0	30.6	19.0	66.7	14.6	5.6	0.0	27.4	2.1	2.8	0.0	13.1
Rape/sexual assault	100.0	97.2	95.2	100.0	83.3	63.9	38.1	92.9	72.9	47.2	28.6	79.8
Robbery	79.2	69.4	33.3	90.5	35.4	30.6	19.0	54.8	22.9	22.2	14.3	44.0
Assault	25.0	5.6	0.0	36.9	8.3	5.6	0.0	15.5	2.1	2.8	0.0	7.1
Aggravated	75.0	55.6	23.8	82.1	33.3	8.3	0.0	42.9	14.6	11.1	0.0	25.0
Simple	31.3	11.1	0.0	42.9	12.5	8.3	0.0	17.9	4.2	2.8	0.0	10.7
Personal theft	100.0	97.2	85.7	100.0	66.7	44.4	38.1	85.7	50.0	36.1	23.8	69.0
Property crime	0.0	0.0	0.0	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Household burglary	14.6	2.8	0.0	23.8	2.1	0.0	0.0	10.7	0.0	0.0	0.0	2.4
Motor vehicle theft	54.2	36.1	28.6	72.6	10.4	11.1	0.0	32.1	2.1	11.1	0.0	15.5
Theft	2.1	0.0	0.0	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Violent crimes involving a weapon	64.6	52.8	23.8	78.6	20.8	16.7	0.0	34.5	12.5	13.9	0.0	23.8
Violent crimes involving a firearm	85.4	75.0	57.1	97.6	50.0	38.9	19.0	73.8	33.3	25.0	4.8	56.0
Violent crimes committed by a stranger	45.8	27.8	0.0	59.5	10.4	8.3	0.0	23.8	2.1	2.8	0.0	9.5
Violent crimes committed by an intimate	85.4	88.9	66.7	98.8	58.3	30.6	9.5	73.8	35.4	8.3	0.0	40.5
Violent crimes committed by other relative	93.8	94.4	90.5	100.0	77.1	55.6	38.1	85.7	56.3	36.1	28.6	69.0
Violent crimes committed by other known offender	54.2	27.8	4.8	70.2	29.2	8.3	0.0	33.3	18.8	5.6	0.0	20.2
Violent crimes occurring during the day	31.3	19.4	0.0	53.6	12.5	8.3	0.0	22.6	4.2	8.3	0.0	15.5
Violent crimes occurring at night	35.4	19.4	4.8	45.2	12.5	8.3	0.0	19.0	4.2	2.8	0.0	14.3
HH crimes occurring during the day	4.2	5.6	0.0	11.9	0.0	0.0	0.0	1.2	0.0	0.0	0.0	1.2
HH crimes occurring at night	4.2	2.8	0.0	7.1	0.0	0.0	0.0	1.2	0.0	0.0	0.0	1.2

Notes: Estimates based on 10 or fewer sample cases, or with relative standard errors greater than 30%, are flagged as unreliable. HH = household.

Table 3.5b. Percentage of victimization rates flagged as unreliable by type of crime: Option 2—crimes reported to police

Type of crime	1-year				3-year				5-year			
	Reg. × pop. size	Reg. × urban-icity	Urban-icity × pop. size	Reg. × pop. size × urban-icity	Reg. × pop. size	Reg. × urban-icity	Urban-icity × pop. size	Reg. × pop. size × urban-icity	Reg. × Pop. Size	Reg. × urban-icity	Urban-icity × pop. size	Reg. × pop. size × urban-icity
All crime	2.1	2.8	0.0	8.3	0.0	0.0	0.0	3.6	0.0	0.0	0.0	0.0
Violent crime	33.3	16.7	9.5	44.0	4.2	8.3	0.0	19.0	2.1	2.8	0.0	13.1
Serious violent crime	70.8	52.8	14.3	79.8	18.8	8.3	0.0	32.1	10.4	8.3	0.0	20.2
Rape/sexual assault	100.0	100.0	100.0	100.0	89.6	83.3	81.0	98.8	85.4	66.7	61.9	96.4
Robbery	89.6	77.8	57.1	96.4	50.0	50.0	23.8	70.2	31.3	33.3	14.3	53.6
Assault	43.8	22.2	4.8	51.2	20.8	8.3	0.0	27.4	4.2	5.6	0.0	14.3
Aggravated	85.4	72.2	52.4	94.0	39.6	16.7	4.8	57.1	22.9	8.3	0.0	33.3
Simple	58.3	38.9	9.5	72.6	27.1	5.6	0.0	31.0	12.5	5.6	0.0	19.0
Personal theft	100.0	100.0	100.0	100.0	97.9	88.9	57.1	98.8	72.9	55.6	47.6	92.9
Property crime	6.3	0.0	0.0	10.7	0.0	0.0	0.0	1.2	0.0	0.0	0.0	0.0
Household burglary	25.0	16.7	0.0	40.5	6.3	0.0	0.0	10.7	2.1	0.0	0.0	3.6
Motor vehicle theft	58.3	44.4	23.8	73.8	14.6	16.7	0.0	39.3	2.1	13.9	0.0	20.2
Theft	8.3	2.8	0.0	16.7	0.0	0.0	0.0	2.4	0.0	0.0	0.0	1.2
Violent crimes involving a weapon	77.1	61.1	38.1	89.3	27.1	11.1	4.8	47.6	16.7	11.1	0.0	29.8
Violent crimes involving a firearm	97.9	86.1	57.1	98.8	58.3	52.8	28.6	83.3	41.7	27.8	4.8	66.7
Violent crimes committed by a stranger	64.6	44.4	14.3	83.3	29.2	13.9	4.8	38.1	6.3	8.3	0.0	19.0
Violent crimes committed by an intimate	93.8	88.9	76.2	100.0	68.8	52.8	28.6	88.1	54.2	13.9	19.0	65.5
Violent crimes committed by other relative	97.9	97.2	95.2	100.0	79.2	66.7	52.4	92.9	66.7	44.4	33.3	78.6
Violent crimes committed by other known offender	79.2	72.2	42.9	91.7	41.7	25.0	4.8	51.2	35.4	22.2	0.0	41.7
Violent crimes occurring during the day	68.8	44.4	19.0	79.8	25.0	16.7	0.0	33.3	14.6	8.3	0.0	22.6
Violent crimes occurring at night	62.5	38.9	9.5	73.8	18.8	16.7	0.0	29.8	14.6	8.3	0.0	23.8
HH crimes occurring during the day	16.7	13.9	0.0	28.6	4.2	0.0	0.0	6.0	0.0	0.0	0.0	2.4
HH crimes occurring at night	10.4	11.1	0.0	21.4	2.1	0.0	0.0	4.8	0.0	0.0	0.0	1.2

Notes: Estimates based on 10 or fewer sample cases, or with relative standard errors greater than 30%, are flagged as unreliable. HH = household.

Table 3.6a. Percentage of victimization rates flagged as unreliable by type of crime: Option 3—crimes reported and not reported to police

Type of crime	1-year				3-year				5-year			
	Reg. × pop. size	Reg. × urban-icity	Urban-icity × pop. size	Reg. × pop. size × urban-icity	Reg. × pop. size	Reg. × urban-icity	Urban-icity × pop. size	Reg. × pop. size × urban-icity	Reg. × Pop. Size	Reg. × urban-icity	Urban-icity × pop. size	Reg. × pop. size × urban-icity
All crime	0.0	0.0	0.0	8.3	0.0	0.0	0.0	5.2	0.0	0.0	0.0	2.1
Violent crime	8.3	8.3	4.2	37.5	0.0	5.6	0.0	15.6	0.0	2.8	0.0	11.5
Serious violent crime	35.4	30.6	25.0	68.8	10.4	5.6	0.0	34.4	2.1	2.8	0.0	18.8
Rape/sexual assault	100.0	97.2	95.8	100.0	83.3	63.9	37.5	94.8	66.7	47.2	33.3	82.3
Robbery	72.9	69.4	41.7	93.8	29.2	30.6	25.0	61.5	20.8	22.2	12.5	53.1
Assault	16.7	5.6	4.2	42.7	2.1	5.6	0.0	18.8	0.0	2.8	0.0	12.5
Aggravated	66.7	55.6	33.3	84.4	22.9	8.3	0.0	47.9	8.3	11.1	4.2	29.2
Simple	20.8	11.1	4.2	50.0	4.2	8.3	0.0	20.8	0.0	2.8	0.0	14.6
Personal theft	100.0	97.2	79.2	100.0	66.7	44.4	58.3	87.5	39.6	36.1	41.7	71.9
Property crime	0.0	0.0	0.0	7.3	0.0	0.0	0.0	3.1	0.0	0.0	0.0	1.0
Household burglary	8.3	2.8	0.0	32.3	0.0	0.0	0.0	15.6	0.0	0.0	0.0	6.3
Motor vehicle theft	45.8	36.1	41.7	75.0	4.2	11.1	0.0	43.8	0.0	11.1	0.0	24.0
Theft	0.0	0.0	0.0	6.3	0.0	0.0	0.0	4.2	0.0	0.0	0.0	2.1
Violent crimes involving a weapon	62.5	52.8	29.2	83.3	14.6	16.7	0.0	40.6	4.2	13.9	0.0	26.0
Violent crimes involving a firearm	81.3	75.0	62.5	95.8	50.0	38.9	37.5	80.2	27.1	25.0	4.2	61.5
Violent crimes committed by a stranger	29.2	27.8	0.0	58.3	6.3	8.3	0.0	31.3	0.0	2.8	0.0	17.7
Violent crimes committed by an intimate	85.4	88.9	62.5	99.0	56.3	30.6	16.7	78.1	29.2	8.3	4.2	50.0
Violent crimes committed by other relative	95.8	94.4	95.8	100.0	75.0	55.6	41.7	87.5	52.1	36.1	12.5	76.0
Violent crimes committed by other known offender	45.8	27.8	16.7	76.0	16.7	8.3	4.2	35.4	12.5	5.6	0.0	22.9
Violent crimes occurring during the day	25.0	19.4	4.2	57.3	6.3	8.3	0.0	27.1	2.1	8.3	0.0	20.8
Violent crimes occurring at night	22.9	19.4	12.5	51.0	8.3	8.3	0.0	27.1	0.0	2.8	0.0	20.8
HH crimes occurring during the day	0.0	5.6	0.0	16.7	0.0	0.0	0.0	4.2	0.0	0.0	0.0	4.2
HH crimes occurring at night	0.0	2.8	0.0	14.6	0.0	0.0	0.0	5.2	0.0	0.0	0.0	4.2

Notes: Estimates based on 10 or fewer sample cases, or with relative standard errors greater than 30%, are flagged as unreliable. HH = household.

Table 3.6b. Percentage of victimization rates flagged as unreliable by type of crime: Option 3—crimes reported to police

Type of crime	1-year				3-year				5-year			
	Reg. × pop. size	Reg. × urban-icity	Urban-icity × pop. size	Reg. × pop. size × urban-icity	Reg. × pop. size	Reg. × urban-icity	Urban-icity × pop. size	Reg. × pop. size × urban-icity	Reg. × Pop. Size	Reg. × urban-icity	Urban-icity × pop. size	Reg. × pop. size × urban-icity
All crime	0.0	2.8	0.0	14.6	0.0	0.0	0.0	6.3	0.0	0.0	0.0	1.0
Violent crime	25.0	16.7	20.8	50.0	4.2	8.3	0.0	28.1	0.0	2.8	0.0	19.8
Serious violent crime	62.5	52.8	29.2	81.3	14.6	8.3	4.2	39.6	6.3	8.3	0.0	27.1
Rape/sexual assault	100.0	100.0	100.0	100.0	97.9	83.3	79.2	100.0	89.6	66.7	62.5	96.9
Robbery	93.8	77.8	70.8	99.0	41.7	50.0	33.3	76.0	25.0	33.3	25.0	60.4
Assault	31.3	22.2	16.7	55.2	6.3	8.3	0.0	31.3	0.0	5.6	0.0	20.8
Aggravated	83.3	72.2	54.2	94.8	37.5	16.7	16.7	61.5	8.3	8.3	4.2	36.5
Simple	50.0	38.9	12.5	79.2	18.8	5.6	0.0	39.6	6.3	5.6	0.0	25.0
Personal theft	100.0	100.0	100.0	100.0	97.9	88.9	70.8	99.0	79.2	55.6	66.7	94.8
Property crime	0.0	0.0	0.0	15.6	0.0	0.0	0.0	5.2	0.0	0.0	0.0	1.0
Household burglary	20.8	16.7	4.2	51.0	0.0	0.0	0.0	14.6	0.0	0.0	0.0	9.4
Motor vehicle theft	54.2	44.4	41.7	77.1	10.4	16.7	0.0	47.9	0.0	13.9	0.0	29.2
Theft	2.1	2.8	0.0	21.9	0.0	0.0	0.0	7.3	0.0	0.0	0.0	5.2
Violent crimes involving a weapon	72.9	61.1	45.8	93.8	16.7	11.1	16.7	54.2	6.3	11.1	0.0	36.5
Violent crimes involving a firearm	100.0	86.1	70.8	99.0	60.4	52.8	41.7	85.4	35.4	27.8	16.7	71.9
Violent crimes committed by a stranger	58.3	44.4	12.5	82.3	16.7	13.9	8.3	42.7	2.1	8.3	0.0	28.1
Violent crimes committed by an intimate	93.8	88.9	83.3	100.0	64.6	52.8	37.5	90.6	47.9	13.9	20.8	75.0
Violent crimes committed by other relative	100.0	97.2	95.8	100.0	79.2	66.7	54.2	95.8	68.8	44.4	33.3	84.4
Violent crimes committed by other known offender	79.2	72.2	29.2	93.8	33.3	25.0	16.7	58.3	20.8	22.2	8.3	45.8
Violent crimes occurring during the day	54.2	44.4	33.3	82.3	18.8	16.7	4.2	43.8	4.2	8.3	0.0	27.1
Violent crimes occurring at night	56.3	38.9	20.8	78.1	12.5	16.7	0.0	34.4	8.3	8.3	0.0	27.1
HH crimes occurring during the day	8.3	13.9	4.2	34.4	0.0	0.0	0.0	12.5	0.0	0.0	0.0	7.3
HH crimes occurring at night	2.1	11.1	0.0	26.0	0.0	0.0	0.0	10.4	0.0	0.0	0.0	7.3

Notes: Estimates based on 10 or fewer sample cases, or with relative standard errors greater than 30%, are flagged as unreliable. HH = household.

3.5 Weight Recommendations

On the basis of the results of the weighting evaluation, it is recommended that uncalibrated (i.e., original) weights be used in the analyses of these four types of generic areas. Although the coverage evaluation demonstrated some differences in the demographic composition of the generic areas and the NCVS weight sums for these demographic groups, there was no evidence that these demographic differences led to substantive changes in key estimates. The calibrated weights also reduced the precision of resulting estimates and would be more difficult for external researchers to replicate. For these reasons, it is preferable to use the original, uncalibrated NCVS weights in the analyses of generic areas based on the three PUF subnational variables.

Because maintaining all levels of the population size variable can lead to low levels of precision, particularly within the Census region × population size × urbanicity generic areas, it is recommended that this variable be collapsed into four levels: not in a place, <100,000–249,999, 250,000–999,999, and 1,000,000+. This collapsing option provided the most precision gains while maintaining reasonably distinct levels from a substantive perspective. Although this collapsing option does not ameliorate all precision concerns for these generic areas, it allows precise estimates to be calculated for the majority of areas and crime types.

When calculating estimates within generic areas, it is recommended that data be pooled across multiple years to ensure reasonable precision levels for key estimates. Depending on the generic areas included in an analysis and the crime type, pooling 3 or 5 years of data often provides adequate precision within generic areas.

SECTION 4. EVALUATION OF GENERALIZED VARIANCE FUNCTIONS IN SUBNATIONAL AREAS

4.1 Overview of Generalized Variance Function Evaluation

The objective of this analysis is to assess whether the GVFs created by the Census Bureau provide variance estimates consistent with direct variance estimation at the subnational level. Although the NVAT currently uses GVFs to calculate variances, GVFs were designed to provide variance estimates at the national level; their accuracy has not been evaluated at the subnational level. If the estimates in a subnational area deviate greatly from the national estimates, then the parameters developed for the national estimates may not be appropriate for the subnational area.

There is sufficient information available on the PUFs to account for the complex survey design and to accurately approximate variances using direct variance estimation. Of the several methods for calculating direct variance estimates, the two most common are Taylor series linearization (TSL) and balanced repeated replication (BRR). For this analysis, TSL was used as the direct variance method, as recommended by the NCVS direct variance evaluation report (Williams et al., 2014).

Both direct and GVF variances were calculated for key NCVS estimates within the four generic area types, and the resulting precision estimates were compared for an evaluation of whether or not the GVFs produced accurate variance estimates at the subnational level. This section presents the results of these comparisons and provides recommendations for calculating variances within generic areas for BJS reports and within the NVAT.

4.2 Comparison of Direct Estimates to GVFs in Subnational Areas

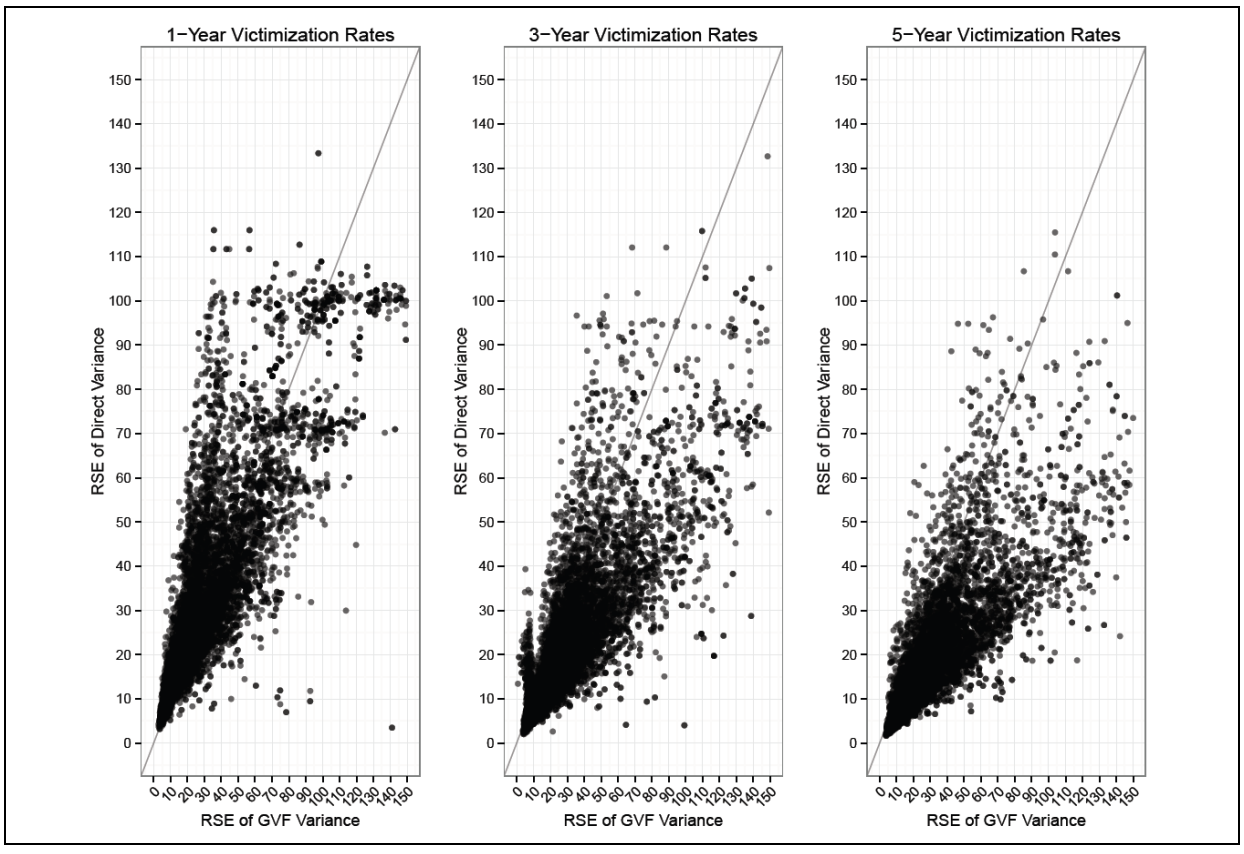
Within each generic area, victimization totals, rates, and their accompanying variance estimates were computed for the same crime types (both overall and reported to police) that were included in the weighting evaluation (listed in ***Section 3.3***). Variance estimates were computed using direct variance estimation and GVFs. GVF estimates were based on the series-adjusted

GVF parameters and were aggregated for pooled-year estimates using Census-provided correlations based on the crime type. The following years were included in the analysis:

- 1-year estimates: 2002, 2008, 2012
- 3-year estimates: 2000–2002, 2006–2008, 2010–2012
- 5-year estimates: 1998–2002, 2004–2008, 2008–2012

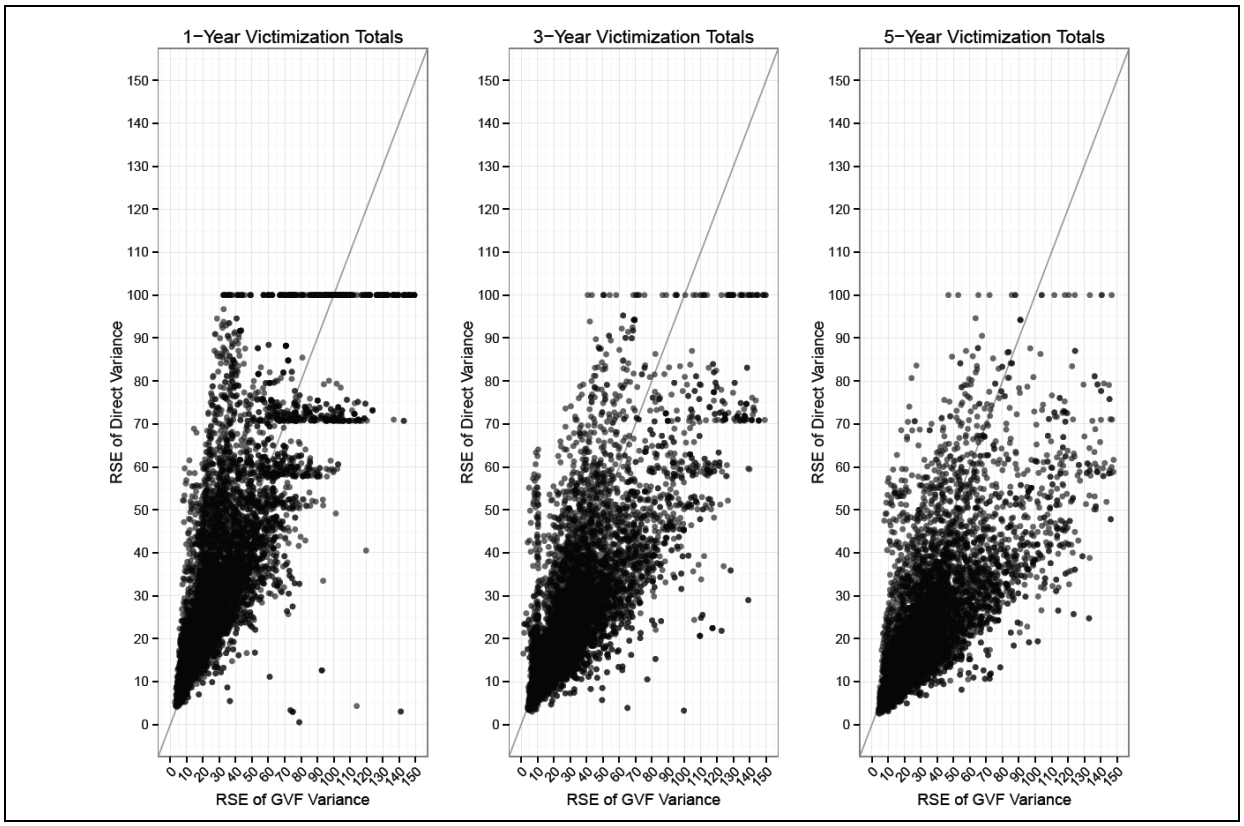
For each number of pooled years, 8,694 estimates were computed for both rates and totals across the analysis years, crime types, and generic area types. The two estimation approaches produce the same point estimates, but variance estimates will be consistent only if the GVFs produce valid estimates at the subnational level. Differences in precision estimates for the two methods may be evaluated by comparing the GVF and direct variance RSEs for rates and totals (***Figures 4.1*** and ***4.2***, respectively) for all generic areas and 1-, 3-, and 5-year periods.

Figure 4.1. Comparison of GVF to direct variance RSEs: 1-, 3-, and 5-year victimization rates



Notes: GVF = generalized variance function; RSE = relative standard error.

Figure 4.2. Comparison of GVF to direct variance RSEs: 1-, 3-, and 5-year victimization totals



Notes: GVF = generalized variance function; RSE = relative standard error.

From ***Figures 4.1*** and ***4.2***, it is clear that the GVFs and direct variances do not track well, given the high levels of deviation from the 45-degree line of equality. For the 1-year victimization rates and totals, the majority of points fall above the 45-degree line, indicating that the GVF approach yielded smaller standard errors than the direct estimation approach. Because the direct estimation approach has been validated with the GVFs at the national level (see Williams et al., 2014) and GVFs were not designed to produce subnational estimates, it is reasonable to assume that the GVF method tends to underestimate the variances of 1-year estimates, whereas the direct variance approach produces accurate variance estimates. Therefore, the use of the GVF for 1-year generic estimates will lead to an increase in the Type I error rate, whereby more comparisons will be deemed significantly different than should be.

The reverse is true for 3- and 5-year victimization rates and totals, where the majority of points fall below the 45-degree line. This indicates that GVF standard errors are larger than direct estimation standard errors, and thus GVFs are overestimating the variances. The use of the

GVF for 3- and 5-year generic estimates will lead to an increase in the Type II error rate, whereby more comparisons will be deemed statistically similar than should be. The same patterns held when concentrating only on reliable estimates (those with GVF and direct variance RSEs of less than 30 percent; results not shown).

Table 4.1 further compares differences across 1-, 3-, and 5-year estimates by displaying the percentage of estimates for which the GVF standard error is less than the direct estimation standard error (i.e., the percentage of estimates for which the GVF is underestimating the variance). Because GVF estimates are dependent upon the GVF parameters, which vary by year, ***Table 4.1*** also splits out these estimates by the analysis year. In addition, ***Table 4.1*** allows for the comparison of estimates across generic area types.

Table 4.1. Percentage of estimates for which GVF standard error is less than direct variance standard error: Differences by year and generic area type

		1-year		3-year		5-year	
Generic area	**Year**	**Totals**	**Rates**	**Totals**	**Rates**	**Totals**	**Rates**
Region × population size × urbanicity	2002	66.8	61.5	43.4	36.5	28.6	16.4
	2008	54.2	51.3	26.2	15.3	12.7	5.4
	2012	48.4	42.8	25.9	13.8	16.8	5.8
Region × population size	2002	61.1	60.3	30.0	28.3	11.3	8.7
	2008	55.6	55.8	16.6	13.2	3.3	2.6
	2012	48.6	48.8	13.5	12.1	8.2	5.6
Region × urbanicity	2002	75.2	66.8	43.3	31.9	29.9	17.8
	2008	61.6	57.6	33.0	15.8	14.7	4.0
	2012	60.3	53.3	30.6	15.9	21.6	6.3
Urbanicity × population size	2002	68.9	66.5	34.8	25.2	21.1	11.8
	2008	62.1	59.9	18.0	6.8	5.6	1.2
	2012	60.2	57.8	18.6	8.7	13.0	3.1

Note: GVF = generalized variance function.

If the GVF and direct estimation approaches provided consistent results, then entries in the table would track right around 50 percent. That is, half the time GVF estimates would track slightly above direct estimates and the other half the time GVF estimates would track slightly below direct estimates. However, ***Table 4.1*** demonstrates the same patterns that were present in ***Figures 4.1*** and ***4.2***: 1-year GVF estimates tend to underestimate variances (percentages are greater than 50 percent), whereas 3- and 5-year estimates tend to overestimate variances

(percentages are less than 50 percent). This result is true for both totals and rates and is independent of the generic area type.

Clear differences can be seen across analysis years, as GVF variance estimates for 2002 are smaller than direct estimates more frequently than those for 2008 and 2012 for 1, 3, and 5 years (i.e., the entries in ***Table 4.1*** are greater for 2002 than for 2008 and 2012). These findings are in line with something that was noted during the 2002 analysis. When GVFs were calculated for the 3-year period of 2000–2002, a total of 146 estimates for rates and totals (84 rates and 62 totals) had negative GVF estimates, and all of them were associated with property crimes. For example, the 2000–2002 estimated rate of motor vehicle theft in Western/rural areas was calculated to be negative on the basis of the GVFs. These negative variance estimates are due to negative GVF parameters in 2000 and 2001 for overall property crime estimates. Although this situation is rare (negative variance estimates were computed for only 2.9 percent of victimization rates and 2.1 percent of victimization totals in the 2000–2002 period), negative variances can be reported when using the GVFs for an estimate type they were not designed to accommodate (e.g., a subnational level estimate). Even though GVF estimates are rarely negative, 2002 GVFs underestimated the true variances more frequently than did the other two analysis years.

Another pattern that is clear from ***Table 4.1*** is that the more years that are pooled, the more GVFs tend to overestimate the true variance. The percentages in ***Table 4.1*** are much lower for 3- than for 1-year and for 5- than for 3-year estimates, indicating that GVFs overestimate variances more the more years that are pooled. This pattern held for all years and all generic area types analyzed.

In addition to assessing differences across years, comparisons were made to assess differences between GVF and direct variance estimates across crime types. ***Tables 4.2a–4.2d*** present the percentage of estimates for totals and rates for which the GVF standard error is less than the direct variance standard error by generic area type, number of years pooled, and type of crime. ***Table 4.2a*** includes the Census region × population size × urbanicity generic areas, ***Table 4.2b*** includes the Census region × population size generic areas, ***Table 4.2c*** includes the Census region × urbanicity generic areas, and ***Table 4.2d*** includes the urbanicity × population size generic areas.

Table 4.2a. Percentage of estimates for which GVF standard error is less than direct variance standard error: Differences by crime type (region × population size × urbanicity)

Type of crime	1-year		3-year		5-year	
	Totals	Rates	Totals	Rates	Totals	Rates
Reported and not reported to police						
All crime	95.2	94.0	78.6	44.0	66.7	27.4
Violent crime	82.1	73.8	59.5	39.3	38.1	14.3
Serious violent crime	53.6	46.4	27.4	20.2	11.9	7.1
Rape/sexual assault	36.9	38.1	25.0	23.8	17.9	14.3
Robbery	35.7	29.8	20.2	17.9	8.3	6.0
Assault	77.4	72.6	58.3	36.9	34.5	9.5
Aggravated	34.5	34.5	31.0	22.6	15.5	10.7
Simple	78.6	70.2	48.8	33.3	33.3	14.3
Personal theft	9.5	8.3	4.8	2.4	2.4	0.0
Property crime	91.7	94.0	61.9	34.5	47.6	20.2
Household burglary	95.2	86.9	61.9	41.7	46.4	28.6
Motor vehicle theft	67.9	66.7	36.9	36.9	20.2	16.7
Theft	91.7	88.1	59.5	32.1	42.9	15.5
Violent crimes involving a weapon	34.5	29.8	22.6	16.7	7.1	4.8
Violent crimes involving a firearm	20.2	20.2	10.7	13.1	4.8	3.6
Violent crimes committed by a stranger	56.0	50.0	33.3	21.4	15.5	4.8
Violent crimes committed by an intimate	56.0	53.6	35.7	29.8	13.1	8.3
Violent crimes committed by other relative	31.0	32.1	16.7	15.5	11.9	13.1
Violent crimes committed by other known offender	64.3	56.0	31.0	20.2	19.0	11.9
Violent crimes occurring during the day	77.4	64.3	41.7	26.2	22.6	11.9
Violent crimes occurring at night	59.5	50.0	36.9	26.2	23.8	10.7
HH crimes occurring during the day	90.5	79.8	42.9	27.4	29.8	9.5
HH crimes occurring at night	85.7	81.0	42.9	26.2	34.5	7.1

(continued)

Table 4.2a. Percentage of estimates for which GVF standard error is less than direct variance standard error: Differences by crime type (region × population size × urbanicity) (continued)

Type of crime	1-year		3-year		5-year	
	Totals	Rates	Totals	Rates	Totals	Rates
Reported to police						
All crime	90.5	89.3	61.9	32.1	42.9	11.9
Violent crime	66.7	59.5	33.3	19.0	25.0	10.7
Serious violent crime	42.9	34.5	14.3	11.9	4.8	6.0
Rape/sexual assault	20.2	21.4	15.5	14.3	8.3	7.1
Robbery	28.6	21.4	14.3	9.5	3.6	4.8
Assault	60.7	53.6	33.3	19.0	21.4	9.5
Aggravated	33.3	34.5	19.0	15.5	6.0	6.0
Simple	56.0	48.8	31.0	17.9	23.8	10.7
Personal theft	6.0	6.0	1.2	0.0	1.2	0.0
Property crime	83.3	75.0	38.1	15.5	23.8	6.0
Household burglary	83.3	79.8	45.2	38.1	28.6	15.5
Motor vehicle theft	65.5	65.5	36.9	36.9	21.4	19.0
Theft	83.3	67.9	35.7	20.2	19.0	2.4
Violent crimes involving a weapon	28.6	28.6	11.9	9.5	1.2	2.4
Violent crimes involving a firearm	14.3	15.5	4.8	6.0	4.8	2.4
Violent crimes committed by a stranger	41.7	41.7	19.0	14.3	11.9	3.6
Violent crimes committed by an intimate	32.1	28.6	26.2	22.6	8.3	4.8
Violent crimes committed by other relative	25.0	23.8	9.5	7.1	7.1	7.1
Violent crimes committed by other known offender	44.0	35.7	14.3	9.5	8.3	6.0
Violent crimes occurring during the day	52.4	41.7	22.6	13.1	8.3	6.0
Violent crimes occurring at night	47.6	42.9	25.0	15.5	14.3	6.0
HH crimes occurring during the day	82.1	73.8	33.3	25.0	13.1	2.4
HH crimes occurring at night	83.3	76.2	31.0	25.0	15.5	3.6

Notes: GVF = generalized variance function; HH = household.

Table 4.2b. Percentage of estimates for which GVF standard error is less than direct variance standard error: Differences by crime type (region × population size)

Type of crime	1-year		3-year		5-year	
	Totals	Rates	Totals	Rates	Totals	Rates
Reported and not reported to police						
All crime	91.7	91.7	64.6	45.8	39.6	20.8
Violent crime	77.1	77.1	31.3	31.3	8.3	8.3
Serious violent crime	50.0	50.0	14.6	12.5	2.1	0.0
Rape/sexual assault	45.8	47.9	27.1	25.0	16.7	14.6
Robbery	39.6	37.5	10.4	14.6	4.2	4.2
Assault	72.9	70.8	37.5	29.2	10.4	6.3
Aggravated	35.4	35.4	27.1	25.0	6.3	4.2
Simple	75.0	75.0	29.2	25.0	14.6	12.5
Personal theft	12.5	14.6	2.1	0.0	2.1	0.0
Property crime	83.3	93.8	31.3	29.2	18.8	8.3
Household burglary	93.8	89.6	43.8	33.3	31.3	22.9
Motor vehicle theft	70.8	77.1	35.4	35.4	12.5	10.4
Theft	83.3	85.4	33.3	29.2	14.6	10.4
Violent crimes involving a weapon	39.6	37.5	16.7	12.5	2.1	2.1
Violent crimes involving a firearm	22.9	22.9	8.3	8.3	4.2	4.2
Violent crimes committed by a stranger	58.3	56.3	20.8	20.8	4.2	2.1
Violent crimes committed by an intimate	56.3	56.3	27.1	27.1	4.2	2.1
Violent crimes committed by other relative	31.3	33.3	16.7	16.7	16.7	18.8
Violent crimes committed by other known offender	58.3	58.3	16.7	14.6	10.4	8.3
Violent crimes occurring during the day	75.0	64.6	16.7	20.8	6.3	4.2
Violent crimes occurring at night	56.3	52.1	22.9	20.8	8.3	6.3
HH crimes occurring during the day	85.4	87.5	20.8	20.8	4.2	2.1
HH crimes occurring at night	75.0	72.9	18.8	16.7	4.2	2.1

(continued)

Table 4.2b. Percentage of estimates for which GVF standard error is less than direct variance standard error: Differences by crime type (region × population size) (continued)

Type of crime	1-year		3-year		5-year	
	Totals	Rates	Totals	Rates	Totals	Rates
Reported to police						
All crime	85.4	87.5	31.3	14.6	12.5	2.1
Violent crime	62.5	64.6	10.4	10.4	4.2	2.1
Serious violent crime	43.8	41.7	8.3	6.3	0.0	0.0
Rape/sexual assault	25.0	27.1	18.8	18.8	10.4	10.4
Robbery	31.3	31.3	10.4	12.5	2.1	2.1
Assault	54.2	56.3	14.6	16.7	4.2	2.1
Aggravated	39.6	39.6	16.7	14.6	2.1	2.1
Simple	58.3	54.2	14.6	14.6	6.3	4.2
Personal theft	6.3	6.3	2.1	0.0	2.1	0.0
Property crime	70.8	72.9	10.4	4.2	0.0	2.1
Household burglary	79.2	79.2	33.3	31.3	12.5	10.4
Motor vehicle theft	72.9	75.0	35.4	35.4	12.5	14.6
Theft	68.8	64.6	14.6	10.4	0.0	0.0
Violent crimes involving a weapon	31.3	31.3	6.3	6.3	2.1	4.2
Violent crimes involving a firearm	14.6	16.7	6.3	6.3	2.1	2.1
Violent crimes committed by a stranger	43.8	45.8	10.4	10.4	4.2	2.1
Violent crimes committed by an intimate	35.4	35.4	20.8	18.8	0.0	0.0
Violent crimes committed by other relative	29.2	25.0	10.4	10.4	10.4	10.4
Violent crimes committed by other known offender	45.8	41.7	6.3	6.3	2.1	2.1
Violent crimes occurring during the day	56.3	54.2	8.3	8.3	6.3	4.2
Violent crimes occurring at night	50.0	52.1	8.3	8.3	6.3	6.3
HH crimes occurring during the day	70.8	66.7	25.0	20.8	0.0	0.0
HH crimes occurring at night	70.8	72.9	25.0	20.8	0.0	0.0

Notes: GVF = generalized variance function; HH = household.

Table 4.2c. Percentage of estimates for which GVF standard error is less than direct variance standard error: Differences by crime type (region × urbanicity)

Type of crime	1-year		3-year		5-year	
	Totals	Rates	Totals	Rates	Totals	Rates
Reported and not reported to police						
All crime	100.0	97.2	91.7	38.9	75.0	27.8
Violent crime	88.9	83.3	55.6	30.6	44.4	16.7
Serious violent crime	55.6	55.6	22.2	11.1	13.9	5.6
Rape/sexual assault	61.1	52.8	36.1	33.3	13.9	11.1
Robbery	38.9	33.3	16.7	11.1	11.1	11.1
Assault	91.7	80.6	61.1	33.3	36.1	16.7
Aggravated	47.2	41.7	27.8	19.4	11.1	8.3
Simple	86.1	80.6	58.3	33.3	41.7	16.7
Personal theft	8.3	5.6	5.6	2.8	2.8	0.0
Property crime	97.2	91.7	75.0	30.6	58.3	25.0
Household burglary	100.0	94.4	69.4	36.1	55.6	19.4
Motor vehicle theft	66.7	66.7	36.1	33.3	13.9	8.3
Theft	100.0	86.1	61.1	25.0	50.0	16.7
Violent crimes involving a weapon	50.0	44.4	27.8	11.1	11.1	2.8
Violent crimes involving a firearm	25.0	25.0	19.4	16.7	2.8	2.8
Violent crimes committed by a stranger	61.1	58.3	33.3	16.7	19.4	5.6
Violent crimes committed by an intimate	77.8	77.8	44.4	36.1	11.1	0.0
Violent crimes committed by other relative	47.2	41.7	30.6	27.8	16.7	13.9
Violent crimes committed by other known offender	75.0	61.1	41.7	22.2	22.2	11.1
Violent crimes occurring during the day	83.3	77.8	47.2	25.0	30.6	13.9
Violent crimes occurring at night	69.4	58.3	41.7	33.3	22.2	5.6
HH crimes occurring during the day	97.2	86.1	44.4	16.7	33.3	11.1
HH crimes occurring at night	97.2	91.7	44.4	19.4	36.1	8.3

(continued)

Table 4.2c. Percentage of estimates for which GVF standard error is less than direct variance standard error: Differences by crime type (region × urbanicity) (continued)

Type of crime	1-year		3-year		5-year	
	Totals	Rates	Totals	Rates	Totals	Rates
Reported to police						
All crime	97.2	97.2	66.7	25.0	47.2	16.7
Violent crime	77.8	72.2	38.9	25.0	25.0	16.7
Serious violent crime	52.8	50.0	19.4	13.9	5.6	8.3
Rape/sexual assault	36.1	33.3	22.2	27.8	8.3	11.1
Robbery	30.6	25.0	16.7	13.9	11.1	8.3
Assault	80.6	69.4	38.9	22.2	30.6	13.9
Aggravated	44.4	33.3	19.4	11.1	5.6	2.8
Simple	66.7	66.7	38.9	25.0	25.0	11.1
Personal theft	5.6	5.6	2.8	2.8	0.0	0.0
Property crime	88.9	69.4	47.2	8.3	30.6	8.3
Household burglary	94.4	91.7	55.6	33.3	38.9	11.1
Motor vehicle theft	58.3	66.7	36.1	33.3	13.9	11.1
Theft	94.4	66.7	33.3	13.9	22.2	5.6
Violent crimes involving a weapon	38.9	33.3	11.1	11.1	0.0	0.0
Violent crimes involving a firearm	22.2	11.1	8.3	8.3	2.8	0.0
Violent crimes committed by a stranger	47.2	47.2	25.0	13.9	11.1	2.8
Violent crimes committed by an intimate	58.3	55.6	27.8	25.0	8.3	0.0
Violent crimes committed by other relative	38.9	38.9	13.9	13.9	13.9	13.9
Violent crimes committed by other known offender	66.7	55.6	13.9	11.1	13.9	13.9
Violent crimes occurring during the day	63.9	63.9	22.2	16.7	13.9	5.6
Violent crimes occurring at night	55.6	44.4	30.6	22.2	19.4	2.8
HH crimes occurring during the day	86.1	63.9	27.8	16.7	13.9	5.6
HH crimes occurring at night	91.7	72.2	30.6	16.7	19.4	2.8

Notes: GVF = generalized variance function; HH = household.

Table 4.2d. Percentage of estimates for which GVF standard error is less than direct variance standard error: Differences by crime type (urbanicity × population size)

Type of crime	1-year		3-year		5-year	
	Totals	Rates	Totals	Rates	Totals	Rates
Reported and not reported to police						
All crime	90.5	85.7	66.7	38.1	57.1	14.3
Violent crime	81.0	76.2	47.6	14.3	28.6	14.3
Serious violent crime	57.1	61.9	9.5	4.8	0.0	0.0
Rape/sexual assault	61.9	61.9	38.1	38.1	19.0	19.0
Robbery	47.6	42.9	14.3	9.5	4.8	4.8
Assault	81.0	81.0	38.1	14.3	19.0	9.5
Aggravated	57.1	52.4	23.8	23.8	9.5	0.0
Simple	90.5	85.7	38.1	19.0	28.6	14.3
Personal theft	19.0	19.0	4.8	0.0	0.0	0.0
Property crime	81.0	90.5	57.1	19.0	47.6	9.5
Household burglary	100.0	95.2	52.4	19.0	33.3	14.3
Motor vehicle theft	76.2	76.2	28.6	28.6	14.3	9.5
Theft	81.0	85.7	42.9	14.3	38.1	9.5
Violent crimes involving a weapon	57.1	52.4	14.3	4.8	0.0	0.0
Violent crimes involving a firearm	33.3	33.3	9.5	9.5	4.8	4.8
Violent crimes committed by a stranger	57.1	52.4	14.3	4.8	4.8	0.0
Violent crimes committed by an intimate	85.7	85.7	23.8	23.8	0.0	0.0
Violent crimes committed by other relative	57.1	57.1	28.6	28.6	14.3	9.5
Violent crimes committed by other known offender	71.4	61.9	28.6	23.8	14.3	14.3
Violent crimes occurring during the day	71.4	66.7	19.0	19.0	19.0	9.5
Violent crimes occurring at night	76.2	57.1	23.8	9.5	4.8	0.0
HH crimes occurring during the day	95.2	90.5	33.3	9.5	23.8	9.5
HH crimes occurring at night	90.5	85.7	33.3	9.5	23.8	4.8

(continued)

Table 4.2d. Percentage of estimates for which GVF standard error is less than direct variance standard error: Differences by crime type (urbanicity × population size) (continued)

Type of crime	1-year		3-year		5-year	
	Totals	Rates	Totals	Rates	Totals	Rates
Reported to police						
All crime	95.2	85.7	57.1	4.8	33.3	4.8
Violent crime	76.2	66.7	23.8	14.3	9.5	0.0
Serious violent crime	47.6	47.6	4.8	4.8	0.0	0.0
Rape/sexual assault	38.1	33.3	38.1	38.1	14.3	14.3
Robbery	28.6	23.8	0.0	4.8	4.8	4.8
Assault	81.0	71.4	23.8	14.3	4.8	0.0
Aggravated	42.9	42.9	14.3	9.5	0.0	0.0
Simple	76.2	76.2	23.8	9.5	14.3	0.0
Personal theft	9.5	9.5	0.0	0.0	0.0	0.0
Property crime	71.4	76.2	28.6	4.8	19.0	4.8
Household burglary	95.2	90.5	28.6	19.0	23.8	4.8
Motor vehicle theft	61.9	61.9	28.6	33.3	14.3	14.3
Theft	61.9	52.4	23.8	9.5	9.5	0.0
Violent crimes involving a weapon	52.4	52.4	4.8	4.8	0.0	0.0
Violent crimes involving a firearm	23.8	19.0	4.8	4.8	0.0	0.0
Violent crimes committed by a stranger	52.4	52.4	14.3	0.0	0.0	0.0
Violent crimes committed by an intimate	52.4	52.4	9.5	9.5	4.8	0.0
Violent crimes committed by other relative	47.6	47.6	19.0	19.0	19.0	19.0
Violent crimes committed by other known offender	47.6	42.9	4.8	4.8	4.8	4.8
Violent crimes occurring during the day	61.9	61.9	14.3	14.3	4.8	4.8
Violent crimes occurring at night	47.6	52.4	9.5	0.0	0.0	0.0
HH crimes occurring during the day	66.7	66.7	19.0	9.5	9.5	0.0
HH crimes occurring at night	76.2	81.0	9.5	4.8	9.5	0.0

Notes: GVF = generalized variance function; HH = household .

These tables show that the GVFs tend to overestimate the variance for some crime types, whereas for others they tend to underestimate the true variance. As previously noted, 1-year GVFs tend to underestimate variances, but this pattern does not hold for all crime types. For example, GVFs tend to overestimate variances for robbery, personal theft, and violent crimes involving a firearm. One example of a crime type for which the GVFs frequently overestimates the variance is personal theft, for which the GVFs are too high for more than 4 out of 5 estimates for all generic area types, any number of pooled years, and for both overall crimes and crimes reported to the police. Although some crime types exhibit clear patterns, the majority of crime types vary in whether they overestimate or underestimate the variance depending on the number of pooled years and the generic area type. ***Tables 4.2a–4.2d*** also support the conclusions made from ***Table 4.1*** that the more years that are pooled, the more the GVFs overestimate variances (as is evident by the decreasing percentages as more years are pooled within each generic area).

4.3 Variance Estimation Recommendations

On the basis of this analysis, it is clear that the direct estimation and GVF approaches do not provide consistent variance estimates at the generic area level. Because GVFs were designed only to produce national estimates and direct estimation has been validated for use on the NCVS previously, GVFs are evidently not accurate at the subnational level. The use of direct variance estimation within generic areas is recommended for all analysis years and formations of generic areas. This recommendation is based on the following observations:

- GVFs do not track well with direct variance estimates. They tend to underestimate the variances of 1-year estimates and overestimate the variances of pooled-year estimates for both totals and rates, regardless of the generic area type.
- Although rare, negative variances can be reported by using GVFs for subnational estimates.
- Differences in the validity of GVF estimates vary widely based on the analytic year and crime type.

Whenever possible (e.g., for BJS publications), it is recommended that analysts use direct variance estimation, rather than GVFs, for calculating variances within generic areas. Because of the inaccuracies in GVF estimates within generic areas, it is recommended that the three-variable generic area type (Census region × population size × urbanicity) not be added to the NVAT.

SECTION 5. FINDINGS AND RECOMMENDATIONS

Table 5.1 summarizes the findings and recommendations of this report.

Table 5.1. Findings and recommendations regarding the coverage and reliability of NCVS public-use file subnational geographic identifiers

Findings	Recommendations
Coverage assessment	
▪ Assessed the coverage of the National Crime Victimization Survey (NCVS) sample in the generic areas formed from crossing the three subnational identifiers available on the NCVS public-use file: Census region, population size, and urbanicity. ▪ Developed comparable Census estimates within the four types of generic areas. ▪ Compared the Census estimates to NCVS weight sums within generic areas for key person- and household-level demographic characteristics for three comparison years: 2002, 2008, and 2012. ▪ Although the estimates tracked well, there were a nonnegligible number of significant differences.	▪ Further evaluation was recommended to assess whether differences in demographic coverage rates translated to substantive differences in point estimates.
Weighting assessment	

▪ The weights of the NCVS sample were calibrated to Census control totals at the generic area level.
▪ Point estimates and estimated precision levels (percent relative standard errors, or RSEs) for the uncalibrated and calibrated weights were compared.
  ▪ Calibrated and uncalibrated estimates tracked well, indicating that there was no evidence of bias resulting from use of the original (i.e., uncalibrated) NCVS weights.
  ▪ Calibrated weights had lower precision levels than uncalibrated weights.
	▪ It was recommended that uncalibrated weights be used for calculations of key estimates within generic areas formed from the public-use file subnational variables.

Variance estimation	
▪ Two methods for calculating design-consistent variances for the NCVS are direct variance estimation and generalized variance functions (GVFs). ▪ Direct variance estimation has been validated to produce accurate estimates for the NCVS.[1] ▪ GVFs were not designed to produce estimates at the subnational level. ▪ GVF estimates do not track well with direct variance estimates, indicating that they will lead to biased variance estimates within generic areas.	▪ It is recommended that direct variance estimation be used within generic areas whenever possible. ▪ Because the NCVS Victimization Analysis Tool (NVAT) produces variance estimates based on GVFs, it is recommended that the generic areas based on the three-variable cross of Census region, population size, and urbanicity not be added to the NVAT.

[1] 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.

SECTION 6. REFERENCES

Planty, M. (2012). Approaches to estimating subnational victimization estimates. In *Federal Committee on Statistical Methodology Proceedings*, pp. 1–10. Retrieved from http://fcsm.sites.usa.gov/files/2014/05/Planty_2012FCSM_I-B.pdf

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.

The Bureau of Justice Statistics of the U.S. Department of Justice is the principal federal agency responsible for measuring crime, criminal victimization, criminal offenders, victims of crime, correlates of crime, and the operation of criminal and civil justice systems at the federal, state, tribal, and local levels. BJS collects, analyzes, and disseminates reliable and valid statistics on crime and justice systems in the United States, supports improvements to state and local criminal justice information systems, and participates with national and international organizations to develop and recommend national standards for justice statistics. William J. Sabol is director.

Office of Justice Programs
Innovation • Partnerships • Safer Neighborhoods
www.ojp.usdoj.gov