November 18, 2010

Evaluating the Effects of Selecting One Person from Each NCVS Housing Unit on the Precision of Victimization Rates

Final Report

Prepared for

Shannan Catalano, Ph.D.
Department of Justice
Bureau of Justice Statistics
810 7th Street, NW
Washington, DC 20531

Prepared by

Vincent G. Iannacchione, M.S.
Bonnie E. Shook-Sa, M.A.S.
RTI International
3040 Cornwallis Road
Research Triangle Park, NC 27709

RTI Project Number 0211889.001.003

EVALUATING THE EFFECTS OF SELECTING ONE PERSON FROM EACH NCVS HOUSING UNIT ON THE PRECISON OF VICTIMIZATION RATES

by Vincent G. Iannacchione and Bonnie E. Shook-Sa

RTI International

November 18, 2010

CONTENTS

SECTION 1
INTRODUCTION

The National Crime Victimization Survey (NCVS) is a survey of the U.S. civilian non-institutionalized population that focuses on personal and property crimes. Personal crimes are those committed against individuals and include rape/sexual assault, robbery, assault, and personal theft. Property crimes are committed against a household and include burglary, theft, and motor vehicle theft. Data collected from the NCVS are used to estimate yearly victimization rates (VRs)[1] and changes in VRs from year to year.

The NCVS is a panel survey in which each sampled housing unit (HU)[2] or housing unit equivalent (HUE)[3] is interviewed once every 6 months over a 3-year period for a total of seven interviews. Currently, everyone 12 years of age or older in a sampled HU/HUE is asked if they were victims of crimes that occurred during the previous 6 months. In response to a recommendation from the National Research Council (2008), the Bureau of Justice Statistics, which administers the NCVS, is considering restricting the sample to one eligible person from each sampled HU/HUE. This change in the NCVS sample design could have far-reaching and unknown implications for response rates, survey costs, and survey estimates and their associated sampling variances.

In a complex sample design like the one used for the NCVS, a number of factors affect the sampling variance of a survey estimate. These include the effects of stratification, clustering, and unequal weighting caused by differential selection and response rates. The net effect these factors have on sampling variances can be measured with a *design effect* or *deff,* which is the ratio of the actual (i.e., design-consistent) variance of an estimate to the variance of a simple random sample of the same size (Kish, 1965). Design effects quantify how the complex design of the sample affects the precision of survey estimates. The larger the design effect, the bigger the sample size required to maintain the desired level of precision.

Restricting the sample to one eligible person per HU/HUE could have counteracting impacts on the design effects of estimates related to personal VRs. On one hand, the design effects would increase as a result of unequal selection probabilities associated with the selection of one person from multi-person HUs and HUEs. Conversely, selecting one person from each sampled HU/HUE might reduce the design effects because the potential correlation among crimes reported by members of the same HU/HUE is avoided. Note that property crimes would not be affected by one person per HU/HUE sampling because they are reported by a single household respondent within each HU/HUE. The household respondent is the housing unit member chosen by the interviewer who appears to be the most knowledgeable about the

---

[1] Victimization rates are measures of the number of criminal victimizations per 1,000 persons or households.

[2] A housing unit is one or more rooms occupied as separate living quarters or intended for occupancy as separate living quarters. The occupants must have direct access from outside the building or through a common hall or entry.

[3] Group quarters are living quarters where residents share common facilities or receive formally authorized care. For the NCVS, group quarters are divided into clusters of four expected persons. These clusters are referred to as housing unit equivalents.

household composition and is able to answer the household screener questions dealing with crimes against the household (U. S. Department of Justice, 2008).

This report presents the results of a simulation study based on the 2008 NCVS public-use database. The specific research question addressed by the simulation is the following:

How does selecting one person per NCVS HU/HUE affect the design effects associated with personal victimization rates?

To answer this question, we focus on the impact of a one-person selection strategy on the precision of survey estimates. We begin the report with a brief description of the current NCVS sampling and weighting methodology. We then describe the selection of the simulation samples that we used to estimate the VRs and the design effects that may accompany a one-person selection strategy. We estimate the number of HUs and HUEs that would be needed to equalize the precision of the current sample design and a one-person design. We end the report by discussing the advantages and disadvantages of a one-person design.

SECTION 2
SUMMARY OF THE NCVS SAMPLING DESIGN

The NCVS uses a stratified, four-stage sampling design to estimate crime victimization rates for the national civilian non-institutionalized population age 12 and older (U.S. Census Bureau, 2009). At the first stage, Primary Sampling Units (PSUs) are demographic areas consisting of large metropolitan areas, counties, or groups of adjacent counties. Large PSUs are included in the sample automatically and each is assigned its own stratum. These PSUs are considered to be self-representing because all of them are selected. The remaining PSUs, called non-self-representing, are combined into strata by grouping PSUs with similar geographic and demographic characteristics, as determined by the decennial census used to design the sample.

At the second stage, each selected PSU is divided into segments (clusters of about four HUs/HUEs each), and a systematic sample of segments is selected. At the third stage, all HUs/HUEs in a sampled segment are selected, and at the fourth stage, all persons age 12 and older are selected from each sampled HU/HUE. This type of sampling design enables the selection of a self-weighting probability sample of eligible persons. That is, prior to any weighting adjustments for nonresponse or noncoverage, each eligible person has the same design weight which is the inverse of the overall probability of being selected.

Because of the complex sampling design used for the NCVS, the usual sample variance that assumes simple random sampling needs to be multiplied by the design effect to approximate the sample variance associated with the complex design. The design effect is a useful metric to examine when comparing alternative sampling designs. In general, the overall design effect ($DEFF_T$) may be approximated as the product of two components: $DEFF_C$, which is attributable to clustering, and $DEFF_W$, which is attributable to differential sampling rates (or unequal weighting). That is,

$$DEFF_T = DEFF_C * DEFF_W .$$

In a four-stage design used for the NCVS, the design effect attributable to clustering can be approximated as:

$$DEFF_c = 1+(\overline{b}_1 - 1)\rho_1 + (\overline{b}_2 - 1)\rho_2 + (\overline{b}_3 - 1)\rho_3$$

where $\overline{b}_1$ is the average number of sampled persons per PSU, and ρ_1 is the intracluster correlation that measures the homogeneity of the characteristic being measured for persons within the PSUs. Similarly, $\overline{b}_2$ is the average number of sampled persons per segment, and ρ_2 is the intracluster correlation for persons within segments. Finally, $\overline{b}_3$ is the average number of sampled persons per household, and ρ_3 is the intracluster correlation for persons within households.[4]

The design effect attributable to differential sampling rates and weighting adjustments for nonresponse or noncoverage (Kish, 1965) can be expressed as:

[4] See Hanson, Hurwitz, and Madow (1953, p. 401) for more details.

$$DEFF_W = \frac{n \sum w_i^2}{(\sum w_i)^2}$$

where W_i is the analysis weight assigned to respondent *i*.

In the proposed one person per HU/HUE design, $DEFF_C$ would be smaller than that associated with the current NCVS design because the third component ($(\bar{b}_3 - 1)\rho_3$) would be zero. However, $DEFF_W$ is likely to be greater because of the unequal weighting caused by the selection of one person from multi-person HUs and HUEs. The combined effect on $DEFF_T$ would depend on the relative decrease of $DEFF_C$ to the increase in $DEFF_W$. By simulating samples under the one-person design, this study determines the combined effects of $DEFF_C$ and $DEFF_W$ on $DEFF_T$ for estimated victimization rates.

SECTION 3
SELECTION OF THE SIMULATION SAMPLES

The 2008 NCVS public-use database contains survey data for 88,700 respondents in 48,111 unique HUs/HUEs for the 2008 calendar year (U. S. Department of Justice, 2008). A respondent is a person who completed the screening interview during one or two quarters in 2008. Among these, 69,007 respondents (77.8%) are in 28,418 HUs/HUEs with 2 or more respondents. The remaining 19,693 respondents (22.2%) are in single-respondent HUs/HUEs. HUs/HUEs with multiple respondents had between 2 and 12 respondents per HU/HUE. Selecting one respondent from each of the 28,418 multi-person HUs/HUEs would reduce the total number of respondents to 48,111 (19,693+28,418), which is 54.2 percent of the 88,700 NCVS respondents in 2008.

The distribution of respondents per HU/HUE is shown by demographic domain in Table 1. Multi-respondent HUs/HUEs tend to have higher concentrations of male respondents, Hispanic respondents, younger respondents, and respondents in rural areas than single-respondent HUs/HUEs.

Table 1
Percentage distribution of NCVS respondents[1] per HU/HU equivalent

Domain	Number of Respondents per HU/HUE				Total Respondents
	1	2	3	4 +	
Overall	22.2%	45.6%	17.7%	14.5%	88,700
Gender					
Male	19.8%	46.4%	18.3%	15.5%	42,045
Female	24.3%	44.8%	17.2%	13.6%	46,655
Race					
White Only	21.3%	46.9%	17.5%	14.3%	73,425
Black Only	30.6%	38.8%	17.5%	13.1%	9,672
Other	19.2%	39.8%	20.4%	20.6%	5,603
Hispanic Origin					
Hispanic	15.3%	36.8%	23.1%	24.9%	11,728
Non-Hispanic[2]	23.3%	46.9%	16.9%	13.0%	76,972

(continued)

Table 1 (continued)
Percentage distribution of NCVS respondents[1] per HU/HU equivalent

Domain	Number of Respondents per HU/HUE				Total Respondents
	1	2	3	4 +	
Age					
12–15	0.0%	14.4%	39.3%	46.3%	5,238
16–19	6.1%	19.9%	30.5%	43.5%	5,204
20–24	21.3%	38.5%	18.8%	21.4%	6,206
25–34	24.0%	55.5%	12.9%	7.6%	13,859
35–49	20.6%	42.7%	21.1%	15.5%	23,306
50–64	24.4%	52.5%	14.8%	8.3%	20,567
65+	34.3%	54.4%	7.9%	3.4%	14,320
Urbanicity					
Urban	23.5%	44.6%	17.4%	14.5%	69,895
Rural	17.5%	49.2%	18.7%	14.6%	18,805

[1]A respondent is a person who completed the NCVS screening interview during one or more quarters in 2008.
[2]Includes 184 persons with unknown Hispanicity.
Source: 2008 NCVS Public Use files.

To account for the demographic fluctuations associated with selecting one respondent per HU/HUE, we selected 1,000 replicated samples from the NCVS public-use database for the 2008 collection year. For each replicate sample we selected one respondent at random from each of the 28,418 HUs/HUEs with 2 or more respondents. Although some respondents completed the screening interview twice during 2008, each respondent's probability of selection was independent of the number of times he/she responded. The one person per HU/HUE sample consisted of respondents from single-respondent households and respondents selected from multi-respondent HUs/HUEs.

We extracted all of the 2008 collection year interviews and incidents for respondents in the one person per HU/HUE sample. The number of records on the one person per HU/HUE person-level file and incident-level file fluctuated from replicate to replicate depending on which respondents were selected within multi-respondent HUs/HUEs because respondents within HUs/HUEs could have provided varying numbers of interviews and incidents. The one person per HU/HUE samples – comprised of interviews and incidents from single-respondent HUs/HUEs and the selection of one respondent from multi-respondent HUs/HUEs – account for an average of 54.2 percent of the interviews and 60.4 percent of the incidents reported in 2008. Table 2 shows the distribution of incidents in the full sample vs. the one-person replicate samples.

Table 2
Distribution of Incidents[1] in Full vs. One-Person Samples

Domain	Full Sample	One-Person Samples[2]		
		Minimum	Median	Maximum
Overall	1,333	746	805	857
Gender				
Male	698	361	403	443
Female	635	364	402	449
Race				
White Only	1,050	577	627	680
Black Only	193	107	124	141
Other	90	40	54	69
Hispanic Origin				
Hispanic	152	67	82	101
Non-Hispanic[3]	1,181	670	722	778
Age				
12-15	170	28	59	80
16-19	153	41	61	81
20-24	172	86	109	132
25-34	260	154	184	212
35-49	333	187	211	238
50-64	195	129	145	163
65+	50	27	36	46
Urbanicity				
Urban	1,124	639	692	738
Rural	209	85	113	137

[1] Includes all non-series rape/sexual assault, robbery, assault, and personal theft incidents.

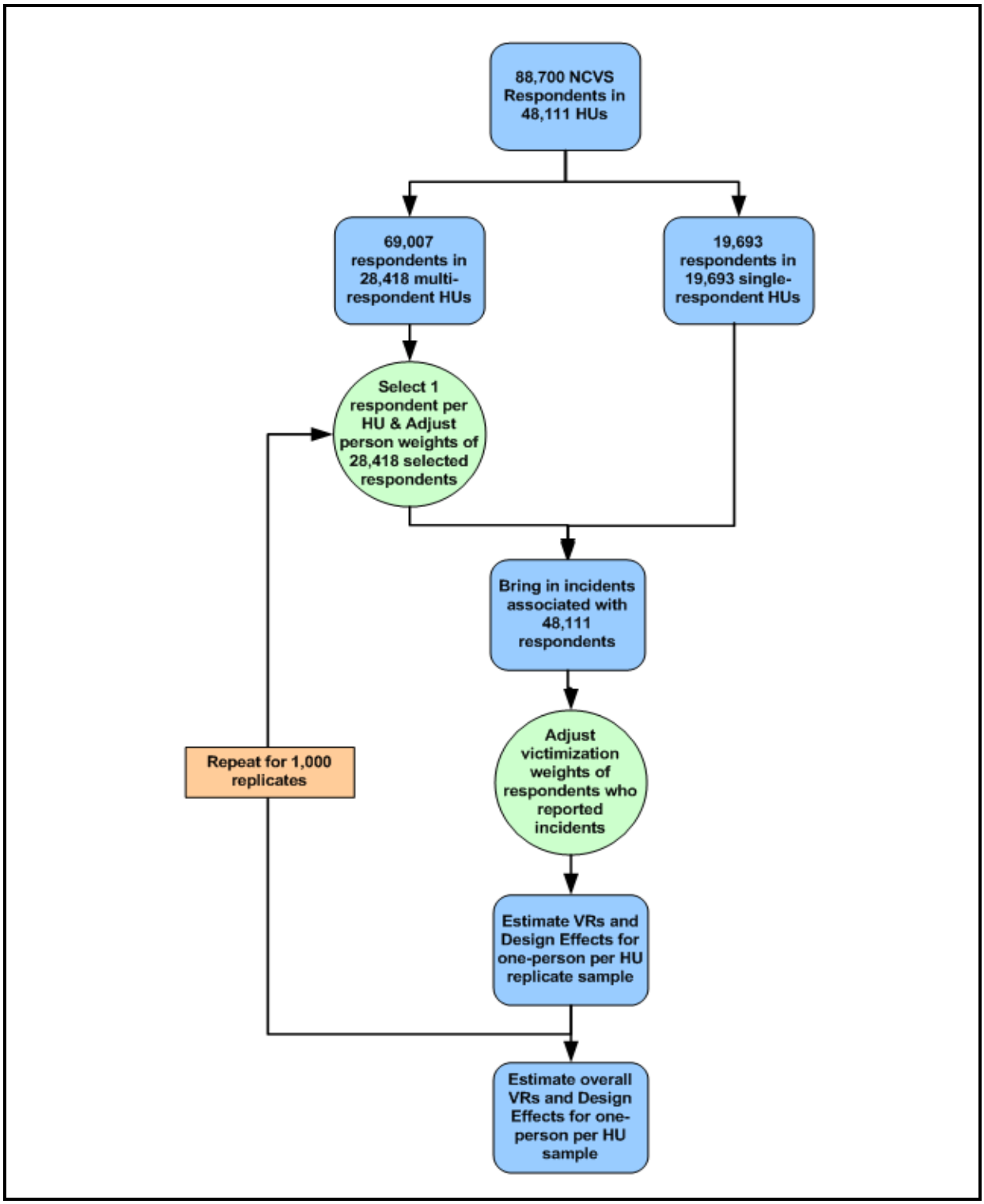
[2] For 1,000 replicate samples.

[3] Includes 184 persons with unknown Hispanicity.

After selecting the one person per HU/HUE person-level sample for each replicate, we applied two adjustment factors to the person weights of respondents in multi-respondent HUs/HUEs to reflect the within-HU/HUE subsampling. First, we adjusted the person-level weights to reflect the respondents' within-HU/HUE probabilities of selection. We then poststratified the weights of respondents representing multi-person HUs/HUEs to full-sample totals for respondents in multi-respondent HUs/HUEs for the 42 domains defined by gender (Male, Female), race (White, Black, Other), and age category (12–15, 16–19, 20–24, 25–34, 35–49, 50–64, 65+). The weights for persons residing in single-respondent HUs/HUEs were not adjusted.

After adjusting the person-level weights for subsampling, we adjusted the weights of victimizations on the incident file by multiplying the victimization weight by the ratio of the new person weight to the original person weight. Because the person weights were not adjusted for respondents residing in single-respondent HUs/HUEs, victimization weights for persons in single-respondent HUs/HUEs did not change with this adjustment. The sampling and weighting process is described in further detail in Appendix A and is outlined in Figure 1 below.

Figure 1
Overview of the sampling, weighting, and estimation process for the one person per HU/HUE samples



VR=Victimization Rate; HU=Housing Unit

SECTION 4
RESULTS

We calculated VRs for the one person per HU/HUE replicate samples using the adjusted victimization weights to estimate the total number of victimizations in the numerator and the estimated number of persons in the denominator. Estimated VRs were calculated by combining victimizations reported by respondents selected from multi-respondent HUs/HUEs with those reported by single-respondent HUs/HUEs. We calculated VRs and their associated sample variances for personal crimes (rape/sexual assault, robbery, assault, and personal theft). VRs for the full sample were calculated using the original victimization weights and person weights for respondents from single and multi-respondent HUs/HUEs. Overall VRs for the one person per HU/HUE sample were obtained by averaging across the VRs obtained in the 1,000 replicate samples. The resulting design effects of VRs were used to compare the precision of estimates obtained from the one-person design to estimates calculated under the existing design.

We generated design effects for VRs for the full sample and each of the replicate one person per HU/HUE samples for rape/sexual assault, robbery, assault, and personal theft for domains of interest. Variance estimates were computed using the SUDAAN software (RTI, 2008) with the pseudostratum code (V2117) and the half-sample code (V2118) as described on page 15 of the 2008 NCVS codebook. We obtained overall design effects for the one-person sample by calculating the mean design effects across the 1,000 replicate samples. The median design effects across the four crimes are shown in Table 3. Detailed tables of design effects associated with individual crimes are presented in Appendix B.

Table 3
Median design effects[1]: Full sample versus one-person sample

	$DEFF_C$[2]		$DEFF_W$[3]		$DEFF_T$[4]	
Domain	Full Sample	One Person[5]	Full Sample	One Person[5]	Full Sample	One Person[5]
Overall	1.26	1.16	1.09	1.53	1.38	1.77
Gender						
Male	1.33	1.21	1.09	1.50	1.45	1.82
Female	1.19	1.07	1.10	1.55	1.31	1.66
Race						
White Only	1.19	1.05	1.08	1.52	1.28	1.59
Black Only	1.39	1.10	1.14	1.61	1.58	1.78
Other	1.20	1.06	1.10	1.43	1.31	1.51
Hispanic Origin						
Hispanic	1.02	0.94	1.08	1.56	1.11	1.46
Non-Hispanic[6]	1.22	1.17	1.09	1.52	1.34	1.79

Table 3 (continued)
Median design effects[1]: Full sample versus one-person sample

Domain	$DEFF_C$[2]		$DEFF_W$[3]		$DEFF_T$[4]	
	Full Sample	One Person[5]	Full Sample	One Person[5]	Full Sample	One Person[5]
Age						
12–15	1.22	1.17	1.07	1.14	1.30	1.34
16–19	1.17	1.06	1.05	1.34	1.24	1.43
20–24	1.16	0.94	1.10	1.48	1.27	1.39
25–34	1.30	1.17	1.10	1.43	1.43	1.67
35–49	0.99	0.88	1.05	1.40	1.04	1.23
50–64	1.09	1.00	1.06	1.33	1.16	1.32
65+	1.16	1.15	1.05	1.21	1.22	1.39
Urbanicity						
Urban	1.29	1.20	1.09	1.55	1.41	1.86
Rural	0.90	0.72	1.09	1.43	0.98	1.02

[1]Median design effects for victimization rates associated with rape/sexual assault, robbery, assault, and personal theft.
[2]Design effect attributable to clustering.
[3]Design effect attributable to unequal weighting.
[4]Overall design effect is the product of *DEFFc* and *DEFFw*.
[5]Median design effects for the one person per household sample are averaged across the 1,000 replicate samples.
[6]Includes 184 persons with unknown Hispanicity.

As expected, the median $DEFF_C$ is lower for the one-person sample than for the full sample because the one person per HU/HUE sample eliminates the intra-household correlation. However, the opposite is true for $DEFF_W$. The unequal weighting effect caused by the unequal probabilities of selection for persons in multi-respondent HUs/HUEs causes $DEFF_W$ to be higher for the one-person sample than for the full sample. This leads to the main finding of this analysis: the loss in precision because of unequal weighting outweighs the gains in precision from eliminating within-household clustering, which leads to an increase in the total design effects ($DEFF_T$). This increase in design effects leads to a loss in precision when estimating VRs with the one person per HU/HUE subsampling.

To determine the stability of the one person per HU/HUE design effects, we calculated the simulation variance and coefficient of variation (CV) for each estimate. The simulation variance and CV of a design effect are defined as:

$$Var(deff) = \frac{1}{R-1}\sum_{r=1}^{R}(deff_r - \overline{deff})^2,$$

where $\overline{deff} = \frac{1}{R}\sum_{r=1}^{R} deff_r$ and R is the number of simulation samples (R = 1,000),

$$CV(deff) = \frac{\sqrt{Var(deff)}}{\overline{deff}}.$$

The CVs associated with the three types of design effects are presented in Table 4. The table shows that $DEFF_C$ is more variable than $DEFF_W$ and accounts for most of the variability in the overall design effect. The stability of the CVs indicates that our conclusions about increased design effects are not subject to excessive random variation. Detailed CV tables for individual VRs are presented in Appendix C.

Table 4
Coefficients of variation of median design effects[1] associated with victimization rates

Domain	$DEFF_C$[2]	$DEFF_W$[3]	$DEFF_T$[4]
Overall	0.20	0.02	0.20
Gender			
Male	0.23	0.03	0.23
Female	0.18	0.03	0.17
Race			
White Only	0.22	0.02	0.22
Black Only	0.22	0.04	0.21
Other	0.36	0.07	0.37
Hispanic	0.28	0.06	0.28
Non-Hispanic[5]	0.20	0.02	0.20
Age			
12–15	0.30	0.02	0.30
16–19	0.27	0.04	0.27
20–24	0.13	0.04	0.12
25–34	0.33	0.07	0.32
35–49	0.27	0.03	0.27
50–64	0.14	0.04	0.14
65+	0.16	0.06	0.20

Table 4 (continued)
Coefficients of variation of median design effects[1] associated with victimization rates

Domain	$DEFF_C$[2]	$DEFF_W$[3]	$DEFF_T$[4]
Urbanicity			
Urban	0.20	0.02	0.20
Rural	0.27	0.04	0.26

[1]Median design effects for victimization rates associated with rape/sexual assault, robbery, assault, and personal theft.
[2]Design effect attributable to clustering.
[3]Design effect attributable to unequal weighting.
[3]Product of *DEFFc* and *DEFFw*.
[4]Design effects for the one person per household sample are averaged across the 1,000 replicate samples.
[5]Includes 184 persons with unknown Hispanicity.

SECTION 5
DISCUSSION

The results of the simulation indicate that subsampling one eligible person from each multi-person HU/HUE is likely to significantly increase the design effects of the VRs. Increased design effects could lead either to increased costs associated with sampling more HUs/HUEs to maintain the current precision of VR estimates or to a loss in precision of estimates within domains of interest.

The simulated selection of one respondent from each of the 28,418 multi-respondent HUs/HUEs that participated in the 2008 NCVS reduced the total number of respondents to 48,111 which is 54.2 percent of the 88,700 NCVS respondents in 2008. The full sample and one person per HU/HUE nominal sample sizes could be equalized by enrolling an additional sample of 88,700-48,111=40,589 one person per HUs/HUEs. Simply equalizing the nominal sample sizes, however, does not consider the increased design effects that are associated with a one person per HU/HUE sample.

Using the one person per HU/HUE approach to achieve the same precision as the 2008 full-sample VR estimates for each of the four types of personal crimes would require one respondent to be enrolled from each of 123,898 HUs/HUEs. This represents an increase of nearly 75,800 participating HUs/HUEs over the 48,111 HUs/HUEs achieved in the 2008 NCVS. Even higher increases would be needed to equalize the precision of the full and one person per HU/HUE samples for all domains of interest.

Our results indicating increased design effects with a one-person design are at odds with those reported by Groves and Heeringa (2006). Their empirical study compared the relative sampling variance associated with selecting one adult in a two-person household to selecting both adults and was conducted as part of the National Comorbidity Survey–Replication (NCS-R), which was a national area probability survey designed to measure the prevalence and severity of mental health disorders in the U.S. household population. The study found that the addition of a second adult respondent in eligible two-person households increased the average sampling variance associated with prevalence rates for mental health diagnoses by 10 to 15 percent. In other words, selecting both persons from a two-person household caused the design effects attributable to intra-household correlation to exceed the reduction in design effects attributable to unequal weighting.

The conclusions of the NCS-R empirical study only apply to a population with two adults per household. As a result, single-adult households, which account for approximately 22 percent of both the NCVS and the NCS-R target populations, are excluded. In general, persons living in single-person households will have a much higher selection probability than those living in multi-person households. For example, a person living in a 2-person household will have a within-household selection probability of 0.5, while a person living in a single-person household will be selected with certainty. Combining data from single-person households with data obtained from multi-person households where subselection was implemented increases the design effects attributable to unequal weighting. This increase in unequal weighting can exceed the reduction in design effects caused by the removal of intra-household correlation, resulting in an increase in the total design effects.

Our simulation study indicates that the unequal weighting that results when multi-person HUs/HUEs are combined with one-person HUs/HUEs more than offsets any reduction in design effects caused by the lack of intra-household correlation in a one person per HH selection. Three important caveats are associated with this analysis.

1. The existing NCVS data for multi-respondent HUs/HUEs may be subject to certain biases for intra-familial crimes such as domestic violence. For example, under the current design a husband who has been interviewed may tell his wife not to report domestic violence. The resulting underreporting could be reduced with a single respondent per HU/HUE design (assuming privacy is maintained in the interview setting) if the wife is selected because she would be the only person interviewed from the HU/HUE.

2. The simulation assumes that the response propensities of NCVS sample members are not significantly affected by the selection of one person per HU/HUE. However, the survey literature suggests that the size of the survey request (intention to interview everyone 12 or older in a HU/HUE vs. one person) may affect response rates (i.e., the greater the burden, the lower the participation rate). Also, attempting to interview everyone in a HU/HUE may result in privacy concerns that cause deliberate concealment of one or more HU/HUE members (Valentine & Valentine, 1971). In addition, a positive (or negative) interview experience for one HU/HUE member may help to gain (or discourage) the cooperation of the other HU/HUE members. This group dynamic would not apply to a single-respondent design.

3. The cost savings associated with interviewing one person instead of two or more persons in a multi-person HU/HUE are not considered. Presumably, these cost savings would offset at least part of the increased cost needed to enroll enough additional HU/HUEs to equalize the precision of the one-person sample with that of the full sample.

Despite these limitations, this research provides an estimate of the loss in statistical precision that would result if the NCVS were to transition to selecting one person per HU/HUE. Although a one-person design could reduce underreporting of certain types of crimes and would reduce the burden on individual HUs/HUEs, the increase in design effects resulting from the one-person design would lead either to higher costs associated with selecting significantly more HU/HUEs or to a loss in statistical precision of NCVS survey estimates.

SECTION 6
REFERENCES

Groves, R. M., & Heeringa, S. G. (2006). Responsive design for household surveys: Tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society A*, 169, part3.

Hanson, M., Hurwitz, W., & Madow, W. (1953). *Sample survey methods and theory, Volume I.* New York: John Wiley & Sons.

Kish, L. (1965). *Survey sampling.* New York: John Wiley & Sons.

National Research Council. (2008). *Surveying victims: Options for conducting the National Crime Victimization Survey.* Panel to review the programs of the Bureau of Justice Statistics. Robert M. Groves and Daniel L. Cork, eds. Committee on National Statistics and Committee on Law and Justice, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Research Triangle Institute (2008). *SUDAAN language manual, release 10.0*. Research Triangle Park, NC: Research Triangle Institute.

United States Census Bureau. (2009). *General description of the 2000 sample design of the National Crime Victimization Survey* (NCVS00SDE01). Internal Census Bureau memorandum from Kathlene Garland, October 22, 2009.

United States Department of Justice, Bureau of Justice Statistics. (2008). *National Crime Victimization Survey, 2008* [Collection Year Record-Type Files] [Computer file]. ICPSR25461-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2009-09-11. doi:10.3886/ICPSR25461

Valentine, C. A., & Valentine, B. L. (1971). *Missing men: A comparative methodological study of underenumeration and related problems*. Washington, DC: U.S. Bureau of the Census.

APPENDIX A:
DETAILED SAMPLING AND WEIGHTING OF SIMULATION SAMPLES

Selecting the One Person per HU/HUE Sample

We selected one respondent at random from each of 28,418 HUs/HUEs with two or more respondents from the NCVS public-use database for the 2008 collection year. Respondents were selected by a simple random sample of all persons in the HU/HUE who responded during any quarter in 2008, and this selection was not influenced by the number of times each person responded. For example, if a HU/HUE contained two persons, one who responded only in the first quarter and the other who responded in quarters 1 and 3, the probability of selection for each respondent was the same (0.5). The one person per HU/HUE sample consisted of respondents from single-respondent households and respondents selected from multi-respondent HUs/HUEs. We repeated the sampling 1,000 times to dampen the fluctuations in the demographic composition of the one person per HU/HUE sample.

For each replicate sample, we extracted all of the 2008 collection year interviews and incidents for respondents in the one person per HU/HUE sample. The number of records on the one person per HU/HUE person-level file and incident-level file fluctuated from replicate to replicate depending on which respondents were selected within multi-respondent HUs/HUEs because respondents within HUs/HUEs could have provided varying numbers of interviews and incidents.

Adjusting the Person Weights for the One-Person per HU/HUE Sample

After obtaining the one person per HU/HUE person-level file for each replicate, we applied two adjustment factors to the person weights for the 2008 collection year (WGTPERCY) of respondents in multi-respondent HUs/HUEs to reflect the within-HU/HUE subsampling. First, we multiplied WGTPERCY for each interview (i) by the total number of respondents in the HU/HUE (NUM_RESP_j) to adjust the weights of selected respondents for within-HU/HUE subsampling:

$$WGTPERCY2_i = WGTPERCY_i * NUM\_RESP_j ; i=1,\ldots,I ; j=1,\ldots,28{,}418$$

where I = the number of interviews associated with one person per HU/HUE respondents from multi-respondent HUs/HUEs,

and 28,418 = the number of multi-respondent HUs/HUEs.

We then poststratified the weights of respondents representing multi-person HUs/HUEs to full sample totals for respondents in multi-respondent HUs/HUEs for the 42 domains defined by gender (Male, Female), race (White, Black, Other), and age category (12–15, 16–19, 20–24, 25–34, 35–49, 50–64, 65+). The weights for persons residing in single-respondent HUs/HUEs were not poststratified. The resulting adjusted person weight (WGTPERCY_NEW) summed to the original person weight (WGTPERCY) totals by race, gender, and age category.

$WGTPERCY\_NEW_i = WGTPERCY_i$ if respondent i comes from a single-respondent HU/HUE,

$WGTPERCY\_NEW_i = WGTPERCY2_i * PS\_Adjust_i$ if respondent i comes from a multi-respondent HU/HUE.

Where PS_Adjust_i is the poststratification adjustment factor associated with the gender, race, and age category domain of interview i=1,…, I, the number of interviews associated with one person per HU/HUE respondents.

The minimum mean poststratification adjustment factor within poststratification domains across all replicates was 0.93; the maximum was 1.09.

Adjusting the Victimization Weights for the One Person per HU/HUE Sample

After adjusting the person-level weights for subsampling, we adjusted the weights of victimizations on the incident file. The victimization weights (WGTVICCY[5]) were adjusted as follows to reflect the one person per HU/HUE subsampling for each replicate sample.

$$WGTVICCY\_NEW_k = WGTVICCY_k * (WGTPERCY\_NEW_i / WGTPERCY_i)$$

if incident k corresponds to a personal crime and person i is part of the one person per HU/HUE subsample where k=1…7,044, which is the total number of incidents on the 2008 incident file. For incidents corresponding to household crimes and incidents for persons not part of the one person per HU/HUE subsample, WGTVICCY_NEW was set to 0. Note that WGTVICCY_NEW=WGTVICCY for persons in single-respondent HUs/HUEs.

Calculating Victimization Rates for the One Person per HU/HUE Sample

Victimization rates for the one person per HU/HUE replicate samples were calculated for crime C in domain D as:

$$VR_{C,D} = \frac{\sum_{k \varepsilon C,D} WGTVICCY\_NEW_k}{\sum_{i \varepsilon D} WGTPERCY_i} * 1000$$

where $\sum_{k \varepsilon C,D} WGTVICCY\_NEW_k$ is the estimated number of victimizations[6] for crime C in domain D, and $\sum_{i \varepsilon D} WGTPERCY_i$ is the estimated number of persons in domain D.[7]

[5] WGTVICCY is used to weight the incident data to count victimizations on a collection-year basis. A personal victimization is a specific criminal act as it affects one victimized person. Estimates of the number of personal victimizations are determined by the number of victims of such acts as reported by the survey respondent. Ordinarily, the number of victimizations is somewhat higher than the number of incidents because more than one person may be victimized during certain incidents, and because details of personal crimes occurring during the course of a commercial crime, as related to the victimized person, are reflected in these estimates.

For example, if we are interested in estimating the robbery VR for males using the one person per HU/HUE sample, we would obtain the sum of WGTVICCY_NEW for all nonseries robberies committed against men in the United States, divide this total by the sum of the person weights (WGTPERCY) for all men, and multiply this ratio by 1,000. Overall VRs were obtained for the one person per HU/HUE sample by taking the average of the VRs across the 1,000 replicate samples.

[6] To be consistent with published NCVS estimates, series victimizations and victimizations occurring outside of the United States are excluded.

[7] We assumed that the estimated number of persons in domain D was a known population parameter with zero variance.

APPENDIX B:
DESIGN EFFECTS FOR THE FULL SAMPLE VS. ONE-PERSON SAMPLE

LIST OF TABLES IN APPENDIX B

Table B.1
Total design effects ($DEFF_T$)

Domain	Rape		Robbery		Assault		Personal Theft	
	Full Sample	One-Person Sample	Full Sample	One-Person Sample	Full Sample	One-Person Sample	Full Sample	One-Person Sample
Overall	1.26	1.76	1.49	1.78	1.49	1.83	1.27	1.46
Gender								
Male	1.81	2.07	1.42	1.77	1.48	1.87	1.36	1.40
Female	1.37	1.70	1.25	1.61	1.38	1.74	0.86	1.26
Race								
White Only	1.25	1.99	1.14	1.48	1.33	1.71	1.30	1.29
Black Only	1.61	1.50	1.83	2.14	1.55	1.91	1.24	1.64
Other	0.58	0.82	1.69	1.66	1.55	1.62	1.08	1.41
Hispanic Origin								
Hispanic	1.01	0.89	1.27	1.61	0.82	1.42	1.21	1.50
Non-Hispanic[1]	1.29	1.86	1.38	1.72	1.51	1.85	1.13	1.24
Age								
12–15	2.07	1.79	1.21	1.25	1.38	1.43	1.23	1.12
16–19	1.14	1.38	1.24	1.48	1.24	1.56	1.76	0.70
20–24	1.83	1.44	1.21	1.39	1.33	1.39	0.96	0.71
25–34	1.53	1.67	1.48	1.72	1.39	1.66	0.84	0.75
35–49	1.03	1.17	1.07	1.13	1.05	1.29	0.89	1.30
50–64	0.92	0.74	1.05	1.24	1.26	1.40	1.50	1.63
65+	1.12	0.81	0.94	1.38	1.33	1.41	1.64	1.65
Urbanicity								
Urban	1.33	1.84	1.55	1.87	1.49	1.88	1.30	1.51
Rural	1.02	1.22	0.62	0.76	1.05	1.17	0.94	0.88

[1]Includes 184 persons with unknown Hispanicity.

Table B.2
Design effects due to clustering ($DEFF_C$)

Domain	Rape		Robbery		Assault		Personal Theft	
	Full Sample	One-Person Sample	Full Sample	One-Person Sample	Full Sample	One-Person Sample	Full Sample	One-Person Sample
Overall	1.15	1.15	1.36	1.17	1.36	1.20	1.16	0.95
Gender								
Male	1.67	1.38	1.31	1.18	1.36	1.25	1.26	0.94
Female	1.25	1.10	1.13	1.04	1.25	1.12	0.79	0.82
Race								
White Only	1.16	1.31	1.06	0.98	1.23	1.13	1.21	0.85
Black Only	1.41	0.93	1.61	1.33	1.36	1.19	1.09	1.02
Other	0.52	0.58	1.54	1.16	1.41	1.13	0.98	0.99
Hispanic Origin								
Hispanic	0.93	0.57	1.17	1.04	0.75	0.91	1.11	0.96
Non-Hispanic[1]	1.18	1.23	1.26	1.13	1.38	1.21	1.03	0.81
Age								
12–15	1.94	1.57	1.13	1.09	1.29	1.25	1.15	0.98
16–19	1.09	1.03	1.17	1.10	1.18	1.17	1.67	0.53
20–24	1.67	0.97	1.11	0.94	1.21	0.94	0.87	0.48
25–34	1.39	1.18	1.34	1.21	1.26	1.17	0.77	0.53
35–49	0.98	0.84	1.02	0.81	1.00	0.92	0.84	0.93
50–64	0.87	0.56	0.99	0.94	1.19	1.06	1.42	1.23
65+	1.06	0.67	0.89	1.14	1.26	1.16	1.55	1.35
Urbanicity								
Urban	1.22	1.19	1.42	1.21	1.36	1.22	1.19	0.98
Rural	0.94	0.86	0.57	0.54	0.97	0.82	0.87	0.62

[1]Includes 184 persons with unknown Hispanicity.

Table B.3
Design effects due to unequal weighting ($DEFF_W$)

Domain	Rape		Robbery		Assault		Personal Theft	
	Full Sample	One-Person Sample	Full Sample	One-Person Sample	Full Sample	One-Person Sample	Full Sample	One-Person Sample
Overall	1.09	1.53	1.09	1.53	1.09	1.53	1.09	1.53
Gender								
Male	1.09	1.50	1.09	1.50	1.09	1.50	1.09	1.50
Female	1.10	1.55	1.10	1.55	1.10	1.55	1.10	1.55
Race								
White Only	1.08	1.52	1.08	1.52	1.08	1.52	1.08	1.52
Black Only	1.14	1.61	1.14	1.61	1.14	1.61	1.14	1.61
Other	1.10	1.43	1.10	1.43	1.10	1.43	1.10	1.43
Hispanic Origin								
Hispanic	1.08	1.56	1.08	1.56	1.08	1.56	1.08	1.56
Non-Hispanic[1]	1.09	1.52	1.09	1.52	1.09	1.52	1.09	1.52
Age								
12–15	1.07	1.14	1.07	1.14	1.07	1.14	1.07	1.14
16–19	1.05	1.34	1.05	1.34	1.05	1.34	1.05	1.34
20–24	1.10	1.48	1.10	1.48	1.10	1.48	1.10	1.48
25–34	1.10	1.43	1.10	1.43	1.10	1.43	1.10	1.43
35–49	1.05	1.40	1.05	1.40	1.05	1.40	1.05	1.40
50–64	1.06	1.33	1.06	1.33	1.06	1.33	1.06	1.33
65+	1.05	1.21	1.05	1.21	1.05	1.21	1.05	1.21
Urbanicity								
Urban	1.09	1.55	1.09	1.55	1.09	1.55	1.09	1.55
Rural	1.09	1.43	1.09	1.43	1.09	1.43	1.09	1.43

[1]Includes 184 persons with unknown Hispanicity.

APPENDIX C:
COEFFICIENTS OF VARIATION FOR DESIGN EFFECTS FOR THE ONE PERSON SAMPLES

LIST OF TABLES IN APPENDIX C

Table C.1
CVs for total design effects ($DEFF_T$)

Domain	Rape	Robbery	Assault	Personal Theft
Overall	0.31	0.16	0.12	0.24
Gender				
Male	0.90	0.18	0.17	0.29
Female	0.20	0.15	0.12	0.26
Race				
White Only	0.47	0.17	0.17	0.26
Black Only	0.18	0.23	0.20	0.58
Other	0.43	0.32	0.26	0.43
Hispanic Origin				
Hispanic	0.31	0.24	0.22	0.58
Non-Hispanic[1]	0.33	0.16	0.12	0.24
Age				
12–15	0.56	0.17	0.24	0.37
16–19	0.29	0.26	0.18	1.00
20–24	0.08	0.13	0.11	0.14
25–34	0.41	0.28	0.21	0.37
35–49	0.53	0.11	0.15	0.38
50–64	0.04	0.17	0.10	0.17
65+	0.05	0.23	0.17	0.22
Urbanicity				
Urban	0.35	0.16	0.13	0.25
Rural	0.44	0.27	0.21	0.24

[1]Includes 184 persons with unknown Hispanicity.

Table C.2
CVs for design effects due to clustering ($DEFF_C$)

Domain	Rape	Robbery	Assault	Personal Theft
Overall	0.31	0.15	0.12	0.24
Gender				
Male	0.90	0.17	0.17	0.28
Female	0.20	0.15	0.12	0.27
Race				
White Only	0.47	0.17	0.17	0.26
Black Only	0.19	0.23	0.21	0.57
Other	0.45	0.28	0.23	0.44
Hispanic Origin				
Hispanic	0.33	0.23	0.20	0.56
Non-Hispanic[1]	0.33	0.16	0.12	0.25
Age				
12–15	0.56	0.17	0.24	0.37
16–19	0.30	0.25	0.18	1.01
20–24	0.10	0.14	0.12	0.16
25–34	0.42	0.29	0.22	0.38
35–49	0.53	0.12	0.16	0.38
50–64	0.07	0.18	0.11	0.18
65+	0.09	0.24	0.13	0.18
Urbanicity				
Urban	0.35	0.15	0.13	0.25
Rural	0.44	0.28	0.21	0.25

[1]Includes 184 persons with unknown Hispanicity.

Table C.3
CVs for design effects due to unequal weighting ($DEFF_W$)

Domain	Rape	Robbery	Assault	Personal Theft
Overall	0.02	0.02	0.02	0.02
Gender				
Male	0.03	0.03	0.03	0.03
Female	0.03	0.03	0.03	0.03
Race				
White Only	0.02	0.02	0.02	0.02
Black Only	0.04	0.04	0.04	0.04
Other	0.07	0.07	0.07	0.06
Hispanic Origin				
Hispanic	0.06	0.06	0.06	0.06
Non-Hispanic[1]	0.02	0.02	0.02	0.02
Age				
12–15	0.02	0.03	0.03	0.02
16–19	0.03	0.04	0.04	0.04
20–24	0.04	0.04	0.04	0.04
25–34	0.07	0.07	0.07	0.07
35–49	0.03	0.03	0.03	0.03
50–64	0.04	0.04	0.04	0.04
65+	0.06	0.05	0.06	0.06
Urbanicity				
Urban	0.02	0.02	0.02	0.02
Rural	0.04	0.04	0.04	0.04

[1]Includes 184 persons with unknown Hispanicity.