

Supporting Statement –

National Crime Victimization Survey Instrument Redesign and Testing Project: Field Test (National Survey of Crime and Safety)

B. Collection of Information Employing Statistical Methods

The NCVSIRTP field test will compare three questionnaire conditions, as shown in Exhibit 1. Condition 1 is the current NCVS questionnaire, interviewer-administered using the same computer-assisted personal interview (CAPI) program and procedures as those used by the Census Bureau. Condition 2 is the redesigned NCVS web-based questionnaire adapted for interviewer administration and run on the interviewers’ laptops. Both Conditions 1 and 2 will include a single interview with a household respondent and each other household member age 12 or older. Conditions 1 (interviewer-administered, current NCVS) and 2 (interviewer-administered, redesigned NCVS) also have a common CAPI household roster interview (Attachment 9). Comparison of data from Conditions 1 and 2 will show the effects of the redesigned instrument, holding mode of interview constant.

Exhibit 1. NCVS-R Field Test Design

	Condition 1 Respondents = 3,000 persons	Condition 2 Respondents = 5,000 persons		Condition 3 Respondents= 4,000 persons			
Instrument	Current NCVS instrument	Redesigned NCVS instrument		Redesigned NCVS instrument			
Mode	In person, telephone	In person, telephone		Household roster in person, then web			
Interview	Interviewer-administered	Interviewer-administered		Self-administered			
Incentive	\$0	\$0		\$0 (N = 1,500)		\$20 (N = 2,500)	
Interleaving	None	Yes (N = 2,500)	No (N = 2,500)	Yes (N = 750)	No (N = 750)	Yes (N = 1,250)	No (N = 1,250)

Condition 3 (self-administered, redesigned NCVS) will begin with the same interviewer-administered, household roster interview as Conditions 1 and 2, with a few content questions asked of the household respondent. This interview will also collect contact information for each adult household member and parents of youth in the household. If the household informant is a parent of youth in the household, he or she will be asked for consent to ask the youth to participate in the survey. If the parent of youth in the household is unavailable at this time, the interviewer will follow-up by telephone to collect consent for all youth in the household. Two months after this household roster interview, adults and youth for whom parental consent has been received will be invited to complete

the redesigned NCVS questionnaire on the web, using their own devices. Non-responders will be recontacted in a variety of ways, including in person, but all interviews will be self-administered, whether respondents used their own devices or the interviewers' laptops during nonresponse follow-up. During the household roster interview, interviewers will provide all households with a nonmonetary incentive (i.e., a magnet) to serve as a reminder of the survey when they are contacted two months later to complete the questionnaire on their own.

Condition 3 is intended to provide data on the performance of a web-based questionnaire relative to two key features of the currently ongoing NCVS. One is the panel design, that is, sampled households in the current NCVS remain in the sample for 3.5 years, and eligible household members are interviewed every 6 months for a total of 7 interviews. In the future, if a web survey was implemented for the NCVS, it would be envisioned that at the first time-in-sample interview, interviewers would visit the household and ask respondents to complete the survey on the web at that time. Once rapport was established at the first interview, respondents would be prompted to complete the survey on their own at second through seventh time-in sample interviews. The intent of the two-month delay for Condition 3 is to simulate how respondents may react at the second through seventh time-in-sample interviews. The primary concern is whether they are willing to complete the survey without an interviewer present. The second key feature being tested in Condition 3 is the effect of self-administration on data quality. Comparison of results across Conditions 2 and 3 will estimate the effects of interview mode for the redesigned instrument.

Two other experiments are embedded in the field test design. Conditions 2 and 3 samples will be split between "interleaved" and "non-interleaved" victimization screeners. In the interleaved version, when a respondent reports a victimization incident, such as a theft, s/he will be asked whether the incident also included some other type of crime, such as an attack or break-in. This approach is intended to reduce duplicate reporting of incidents in response to different victimization probes, and to streamline the interview overall. However, there is concern that interleaving could discourage reporting of other incidents, an effect known as "motivated misreporting."¹ The experiment will assess the effect of interleaving on interview length (burden) and victimization rates. The final experiment will test the effects of a promised \$20 gift card on web survey completion rates and data quality. A portion of the Condition 3 sample will be promised a \$20 gift card upon completion of the survey; the remainder will have no incentive. This experiment will test whether respondents are more willing to complete the NCVS instrument, including all crime incident reports, if they are offered an incentive. There is some evidence that an incentive will significantly increase both the overall response rate and the proportion who complete all required incident forms.²

¹ Eckman, S., F. Kreuter, A. Kirchner, A. Jäckle, R. Tourangeau, and S. Presser (2014), "Assessing the Mechanisms of Misreporting to Filter Questions in Surveys," *Public Opinion Quarterly*, 78,721-733; Back, R.L. and S. Eckman (2018) "Motivated Misreporting in Web Panels" *Journal of Survey Statistics and Methodology* 6: 418 - 430.

² Cantor, D., Townsend, R., and Caparosa, A. (May, 2017). *How much does a promise of \$5 Gift Card Buy for a Web Survey of College Students? Probably more than you think!* (panel presentation presenter). Annual Meeting of the

The field test will also include an experimental comparison of two different formats for the invitation letters sent to households (Conditions 1, 2, and 3) prior to the first contact by an interviewer to conduct the household enumeration interview. In addition, for Condition 3 (self-administered, redesigned NCVS), the same experimental comparison of the two formats for the letters will be sent to individual household members approximately two months after completing the household roster interview inviting the person to complete the self-administered web survey (victimization screener and any applicable Crime Incident Report (CIR)).

Approximately half of all households in each condition will receive an advanced letter prior to the household roster interview that uses a traditional letter format with text in paragraph form (see Attachments 4a, 10a, and 10b). The other half will receive an advanced letter that makes use of icons and a question-answer format rather than a paragraph format (see Attachments 4b, 10c, and 10d). Both versions will include the same content. Only the presentation or format of the information will vary. For Condition 3, the invitation letters sent after the household roster interview to invite individual household members to complete the self-administered web survey will also manipulate the format of the content between a traditional paragraph layout and icons with questions and answers (see Attachments 5a and 5b). All eligible respondents within a household will receive the same format, but half of the households within a format at the household roster level will receive the opposite format for the individual invitation letter. With this design, it is possible to evaluate any interaction effects of the format across the two points of contact.

1. Universe and Respondent Selection

The potential universe for the NCVSIRTP field test is all persons age 12 or older living in households in the 48 contiguous States and the District of Columbia. Persons living in Alaska and Hawaii and those living in group quarters are excluded from the universe for operational efficiency and cost. The field test will employ a stratified three-stage sample design: (1) selection of primary sampling units (PSUs), which are individual counties or groups of counties; (2) selection of secondary sampling units (SSUs), which are census tracts or groups of census tracts within sampled PSUs; and (3) selection of households within sampled SSUs.

The probabilities of selection at each stage will be designed to yield an approximately equal probability sample of households, while attaining the target sample sizes for the experimental treatments and yielding approximately uniform sample sizes across PSUs (with the exception of PSUs selected with certainty). These objectives are achieved by sampling with probabilities proportionate to size (PPS) at the first (PSU) and second (SSU) stages, and then sampling with equal probabilities (within SSUs) at the final (household)

American Association for Public Opinion Research, New Orleans, LA.; Cantor, D. and D. Williams (2013) Assessing Interactive Voice Response for the National Crime Victimization Survey. Final Report prepared for the Bureau of Justice Statistics. Table 4-3. <https://www.bjs.gov/content/pub/pdf/Assessing%20IVR%20for%20the%20NCVS.pdf>

stage. As with the NCVS, there is no sampling within households; all household members aged 12 and older are selected with certainty. This approach will result in all sampled individuals having approximately the same probability of selection.

Sample Size

For the field test, the total target achieved sample size is 12,000 completed interviews with sampled individuals (adults and youth). One key set of analyses will focus on comparing victimization rates across treatments. Table 3 shows, for the planned sample sizes given in Exhibit 1 above, the power for key comparisons of victimization rates. In this table, I(1) is the baseline rate (the expected rate based on the current NCVS design) and I(2) is the rate for the comparison group. The first set of columns covers the following main effect comparisons:

- Current vs redesigned instrument (Condition 1 vs. Condition 2);
- Mode of interview in redesigned instrument (Condition 2, interviewer-administered, vs. Condition 3, web self-administered); and
- Interleaving vs. non-interleaving (Conditions 2 and 3 in each arm).

The second set of columns covers the incentive comparison, entirely within Condition 3.

Table 3. Power for main experimental comparisons for field test (two-tailed at $p < .05$) by type of crime and victimization rate (12 month reference period)

	I(1)	I(2)*	Power assessing instrumentation ¹		Power assessing mode of administration ²		Power assessing interleaving ³		Power assessing incentive ⁴	
			3000	5000	5000	4000	4500	4500	1500	2500
Person level (n=)										
Total violent crimes	0.032	0.039	24%		27%		27%		14%	
		0.048	84%		88%		89%		56%	
		0.065	100%		100%		100%		98%	
		0.097	100%		100%		100%		100%	
		0.129	100%		100%		100%		100%	
Rape or sexual assault	0.002	0.003	5%		5%		6%		4%	
		0.004	12%		13%		13%		8%	
		0.007	75%		74%		73%		46%	
		0.012	99%		99%		99%		88%	
		0.024	100%		100%		100%		100%	
Serious violent crime	0.012	0.015	12%		13%		13%		8%	
		0.019	45%		48%		50%		25%	
		0.025	92%		94%		95%		67%	
		0.037	100%		100%		100%		99%	
		0.050	100%		100%		100%		100%	
Assault	0.025	0.030	19%		22%		22%		12%	
		0.038	74%		79%		80%		46%	

	I(1)	I(2)*	Power assessing instrumentation ¹		Power assessing mode of administration ²		Power assessing interleaving ³		Power assessing incentive ⁴	
		0.050	100%		100%		100%		93%	
		0.076	100%		100%		100%		100%	
		0.101	100%		100%		100%		100%	
Household level (n=)			1833	3055	3055	2253	2654	2654	890	1363
Total property crimes	0.185	0.222	91%		94%		95%		64%	
		0.278	100%		100%		100%		99%	
		0.371	100%		100%		100%		100%	
		0.556	100%		100%		100%		100%	
		0.742	100%		100%		100%		100%	
Burglary	0.035	0.042	25%		28%		29%		15%	
		0.052	68%		70%		72%		38%	
		0.070	99%		99%		100%		87%	
		0.105	100%		100%		100%		100%	
		0.139	100%		100%		100%		100%	
Theft	0.144	0.173	80%		86%		86%		51%	
		0.216	100%		100%		100%		95%	
		0.288	100%		100%		100%		100%	
		0.432	100%		100%		100%		100%	
		0.577	100%		100%		100%		100%	

*I(1) = Rates expected based on 6-month unbounded NCVS data. I(2) = increase in rate from experimental treatment. For all crimes except rape and sexual assault, the expected increase is 1.2, 1.5, 2, 3 and 4 times NCVS rates. For rape and sexual assault, the assumptions are 1.2, 1.5, 2, 5 and 10 times the NCVS rate.

¹Condition 1 vs. Condition 2

²Condition 2 vs. Condition 3

³(Conditions 2 and 3, Interleaved) vs. (Conditions 2 and 3, Not interleaved)

⁴Condition 3, \$0 vs. Condition 3, \$20

Given the target number of completed interviews, the number of addresses to be sampled was determined by inflating the number of completes to account for expected address ineligibility, expected nonresponse, and expected household composition. Specifically, the assumptions used in these computations are as follows:

1. Percentage of addresses that are occupied: 88%;
2. Percentage of households completing the roster: 45% with no incentive; 50% with incentive (Condition 3 only);
3. Percentage of households with youth age 12-17: 17%;
4. Average number of adults in households without rostered youth: 1.97;
5. Average number of adults in households with rostered youth: 2.31;
6. Average number of youth per household with rostered youth: 1.35;
7. Adult victimization screener completion rate, Conditions 1 and 2: 90% for household respondent, 50% for others;
8. Adult web victimization screener completion rate, Condition 3: 40% (household

- respondent) or 20% (others) with no incentive; 50% (household respondent) or 30% (others) with incentive;
9. Parent consent rate: 60% (Conditions 1 and 2, and Condition 3 with incentive), 50% (Condition 3, no incentive); and
 10. Youth Screener completion rate: 40% (Conditions 1 and 2), 30% (Condition 3 with incentive), 20% (Condition 3, no incentive).

Assumptions 3-6 above are derived from the NCVS, while the other assumptions are estimates based on other recent household- or person-level surveys Westat has conducted. The response rate estimates are relatively conservative, to increase the likelihood of meeting the sample yield targets.

These assumptions were applied to compute the total number of sampled addresses needed to yield the target sample sizes given in Exhibit 1 above for each condition. The relevant breakdowns are given in Tables 4a and 4b. Under these assumptions, the expected overall number of addresses needed for the field test is 24,016. A 40% reserve sample will also be selected if the response rate assumptions are not met. Given field staffing considerations, the largest number of PSUs that can be worked efficiently is approximately 60.

Table 4a. Key breakdowns of sample sizes and associated assumptions, Conditions 1 and 2

	Rate Assumptions	Count	
		Condition 1	Condition 2
Initial sample (addresses)		5,252	8,755
Non vacant	0.88	4,622	7,704
Roster completed	0.45	2,080	3,467
Rostered HHs without youth	0.83	1,726	2,878
Rostered HHs with youth	0.17	354	589
Adults in rostered HHs without youth	1.97	3,401	5,669
Adults in rostered HHs with youth	2.31	817	1,361
Youths in rostered HHs with youth	1.35	477	796
Avg. HH size among rostered HHs Avg. adults	2.2573 2.04		
Adult screener completes (HHR Other)	0.9 0.5	2,949	4,916
Parent consent	0.6	286	477
Youth screener completes	0.4	115	191
Total screener completes		3,064	5,107
Victims (Adult Youth)	0.21 0.18	640	1,067
Non-victims		2,424	4,040
Victims with completed interview	0.9	576	960
Non-victims with completed interview	1.0	2,424	4,040
Total completed interviews with individuals		3,000	5,000

Table 4b. Key breakdowns of sample sizes and associated assumptions, Condition 3

	Rate Assumptions	Count	
		Incentive	No Incentive
Initial sample (addresses)		5,207	4,802
Non vacant	0.88	4,582	4,226
Roster completed (Incentive None)	0.45 .40	2,062	1,690
Rostered HHs without youth	0.83	1,711	1,403
Rostered HHs with youth	0.17	351	287
Adults in rostered HHs without youth	1.97	3,371	2,764
Adults in rostered HHs with youth	2.31	810	664
Youths in rostered HHs with youth	1.35	473	388
Avg HH size among rostered HHs Avg adults	2.2573 2.04		
Adult HHR screener complete (Incentive None)	0.65 0.52	1,340	879
Other adult screener complete (Incentive None)	0.5 0.35	1,060	608
Parent consent	0.6 0.5	284	194
Youth screener completes	0.48 0.32	136	62
Total screener completes		2,536	1,549
Victims		533	327
Adult	0.21	508	316
Youth	0.18	24	11
Non-victims		2,023	1,239
Victims with completed survey		477	261
Adult (Incentive None)	0.9 0.8	458	253
Youth (Incentive None)	0.8 0.7	19	7.82
Non-victims with completed survey		2,023	1,239
Total completed surveys		2,500	1,500

Sample Frame

The development of the PSU sampling frame will begin with a county-level file containing estimates from the 2012-2016 five-year American Community Survey (ACS) summary tabulations. (Note that independent cities are included in their surrounding counties' estimates in these tabulations, so it is not necessary to take special steps to include these.) For operational efficiency, counties in Alaska and Hawaii will be excluded from this file. Each county-level record will include the following ACS estimates:

- Total population and total household counts;
- Per capita income in the past 12 months (in 2016 inflation-adjusted dollars);
- Household counts by tenure (owned/rented);
- Household counts by Hispanic origin of householder;
- Population by year householder moved into unit;

- Gross rent of renter-occupied housing units (HUs);
- Value of housing unit for owner-occupied HUs;
- Tenure by household income in the past 12 months (in 2016 inflation-adjusted dollars); and
- Tenure by number of units in structure.

In addition to these estimates, the file will contain county-level urban and rural population counts provided in the “Census Rurality Level: 2010” table available via a link on the Census Bureau website (the “county look-up table” referenced in the press release available at <https://www.census.gov/newsroom/press-releases/2016/cb16-210.html>).

Stage 1. Defining and Selecting PSUs

For the NCVSIRTP field test, the PSU measure of size (MOS) will be the five-year 2012-2016 ACS estimate of the total number of households in the PSU. The county-level MOS will be checked against the minimum MOS, and counties with a MOS below the minimum will be combined with other counties as described later to form PSUs. The PSU MOS is the sum of the county MOS across the counties comprising the PSU.

PSUs will be stratified as described below, and one PSU will be selected from each stratum; thus, a total of 60 strata will be formed. The average stratum size (in terms of total MOS) will be computed by dividing the sum of the MOS of all counties by 60, and any county with MOS exceeding 75% of that total will be included in the sample with certainty. If a county exceeds 150% of that total but is less than 225%, that county will be a multi-hit certainty viewed as essentially comprising two PSUs. If a county exceeds 225% of that total but is less than 300%, that county will be a multi-hit certainty viewed as essentially comprising three PSUs. The designation of multi-hit certainty PSUs limits the variation in households’ probabilities of selection.

Counties not meeting the minimum MOS criterion will be combined with adjacent counties, respecting Census division boundaries, with consideration to the maximum point-to-point distance within the combined unit, until the minimum size criterion is met. This process will result in two types of PSUs: single counties; and two or more contiguous counties within the same Census division.

The noncertainty PSUs will be explicitly stratified first by Census division, then by the percentage urban (based on the 2010 census urban and rural population counts). Within each Census division, major strata will be formed by grouping PSUs from the sorted list to create roughly equal-sized groups. Within the major strata, PSUs will be examined with respect to the following characteristics:

- HUs with a Hispanic or Latino head of household;
- Population in same house as previous five years;
- Occupied HUs with an annual income less than \$15,000;

- Renter-occupied HUs in 5 or more unit structure;
- Owner-occupied HUs with a value less than or equal to \$70,000; and
- Renter-occupied HUs with monthly rent less than \$450.

The characteristic with the greatest variation among PSUs in the major stratum will be used to substratify. The substratifiers may differ among major strata. For the substratifier, cutpoints will be selected in a manner that aims to minimize the variation in the sizes (total MOS) of final strata. The stratification will ensure appropriate sample representation with respect to the characteristics used, and improve the precision of survey estimates associated with these characteristics. However, the allocation of sample to the strata will be done proportional to stratum size (i.e., no oversampling is planned).

Within each noncertainty stratum, one PSU will be selected with probability proportional to the PSU MOS.

Stage 2. Preparing Frames and Sampling within PSUs

Following the selection of PSUs, the next stage of selection is the selection of secondary sampling units (SSUs), which will comprise Census tracts or groups of Census tracts. Within a sampled PSU, Census tracts will be combined as necessary to form SSUs of sufficient size (in terms of MOS).

Sampling frame for SSU selection. The development of the SSU sampling frame will begin with a Census tract-level file containing estimates of total population and total number of households from the 2012-2016 five-year ACS summary tabulations. The SSU MOS will be an estimate (the five-year 2012-2016 ACS estimate) of the total number of households in the SSU. The tract-level MOS will be checked against the minimum MOS, and tracts with MOS below the minimum will be combined with other tracts to form SSUs. The SSU MOS is the sum of the tract MOS across the tracts comprising the SSU.

Computation of the minimum SSU MOS. Balancing the desire for a large number of sampled SSUs within each PSU in order to limit the effect of clustering against the desire for a small number of SSUs within each PSU in order to limit the within-PSU travel-related costs, the number of SSUs per PSU to be selected is 20, with one exception. For multi-hit certainty, the number of SSUs to be selected is 20 times the number of hits.

With a total target of 24,016 sampled households, adding a 40% reserve sample results in a total target (with reserve) of 33,622 households. With 60 PSUs, this equates to an average of 560 sampled households per PSU. Since a total of 20 SSUs will be sampled within each PSU, this results in an expected average of about 28 sampled households per SSU.

These rough calculations shed some light on the minimum size of an SSU. However, for a

more statistically efficient sample, the target sample sizes within an SSU (and, thus, the minimum SSU size) will vary among PSUs proportionate to the first-stage stratum sizes. Specifically, the minimum SSU MOS will be set equal to $k M_h / 20$, where M_h is the total MOS for stratum h (i.e., summed across all PSUs in stratum h) and k is the overall sampling rate, i.e., $k = 33622 / \sum_h M_h$. This minimum MOS ensures that SSUs will be large enough to yield an approximately equal-probability sample of households while keeping the workload within PSUs relatively fixed (for operational efficiency).

Formation of SSUs. Tracts with MOS less than the desired minimum will be combined with other tracts within the same PSU to form SSUs. Where needed, tracts will be combined with numerically adjacent tracts, i.e., with tracts that fall immediately above or below the small tract in a list sorted by tract number. Although this procedure will not always result in geographically contiguous combinations of tracts, it is expected that the number of such “split SSUs” will be small. Additionally, any such splitting would be expected to have a generally positive effect statistically (in that the clustering effects will be reduced); the drawback of having a split SSU is strictly operational. This process will result in two types of SSUs: single tracts; and two or more (often, but not necessarily, contiguous) tracts within the same SSU.

Sampling SSUs: Identification of certainty SSUs. SSUs having a MOS at least as large as the target sampling interval will be selected with certainty, and the expected number of hits associated with each such SSU will be calculated. The expected number of hits for certainty SSU j within PSU i (in stratum h) is

$$p_{hij} = \frac{20 C_{hi} M_{hij}}{\sum_{j \in i} M_{hij}},$$

where C_{hi} is the number of hits at the PSU level (which is equal to 1 except for multi-hit certainty PSUs).

When $p_{hij} \geq 1$, SSU j within PSU i will be identified as a certainty.

Sampling SSUs: Number of noncertainty SSUs. The target number of the noncertainty SSUs to be selected in PSU i , b'_i , will be determined by subtracting from the total target number of SSUs the total number of certainty SSU hits within PSU i , and then rounding up the result, as follows:

$$b'_i = \text{ceiling} \left(20 C_{hi} - \sum_{j \in i_c} p_{hij} \right),$$

where i_c denotes the set of certainty SSUs in PSU i . The use of the “ceiling” function, which rounds up the argument, ensures that there are no fewer than 20 expected hits in each

PSU among the certainty and noncertainty SSUs. (In the case of multi-hit certainty PSUs, in this computation, 20 will be replaced by $20 C_{hi}$.)

Sampling SSUs: Selection of noncertainty SSUs. Before sampling, the noncertainty SSUs will be subject to a serpentine sort within each PSU, resulting in a geographically-based implicit stratification. The SSUs will then be systematically sampled with probabilities proportionate to the SSU MOS.

Stage 3. Selection of Households

Sampling frame for selection of households. The sampling frame for selecting households will be the address-based sampling (ABS) frame maintained by Marketing Systems Group (MSG). Within each sampled SSU, addresses on MSG's ABS frame that geocode to within the SSU's boundaries will be included in the sampling frame. MSG's ABS frame originates from the U.S. Postal Service (USPS) Computerized Delivery Sequence file (CDS), and is updated monthly. Although the MSG frame includes all types of addresses, in developing the sampling frame for address sampling for the field test, only locatable (i.e., city-style) residential addresses will be retained; that is, addresses that do not correspond to a physical location (such as P.O. Box addresses) will be excluded. The residential ABS frame also generally excludes group quarters addresses.

Selecting the sample of households. Addresses will be sampled systematically based on a geographic sort within each sampled SSU. A target total sample size (including reserve) of 33,622 addresses will be selected, with equal probabilities within each SSU. The within-SSU sampling rates will be set equal to the overall sampling rate (the k specified in Section 4.2) divided by the unconditional probability of selection of the SSU. The reserve sample will be obtained by systematically sampling 40% of the full (primary plus reserve) sample, sorting the full sample in its original order of selection.

The sample of addresses within each PSU will be randomly assigned to the three experimental conditions (proportional to the target sample sizes for each condition), and then, independently, to treatments within each condition (e.g., the incentive groups for Condition 3). During data collection, the sample will be monitored to determine whether any assumptions affecting yield are falling substantially short of expectations, indicating a potential shortfall in the numbers of completes. If necessary, reserve sample (either the entire reserve or a random subsample within each PSU) will be released.

Considerations for identifying sampled households during fieldwork

With the use of the ABS frame for sampling households (through their addresses), there are a few special considerations.

Geocoding error. Each address in the ABS frame is associated with one and only one

census tract, based on the latitude/longitude assigned to the address through a geocoding process. Occasionally, error in the geocoding of addresses results in an address being assigned to a location in a census tract different from its actual location. In order to maximize coverage, even if an address is truly located in a tract outside the SSU from which it was sampled, that address will still be retained in the sample (as eligible).

Drop points. Drop point addresses are single delivery points that service multiple residences; there is nothing in the address (such as an apartment number) to distinguish the residences. These can be identified through a variable available on the ABS frame, and are generally located in multi-unit structures. Following the sampling of addresses, a list of all drop point addresses selected in to the sample will be generated. Interviewers will be given instructions as to how to enumerate units and sample at the drop point addresses. The field test survey weights will include a factor to appropriately account for adjustments to the probability of selection of households at each drop point address.

Households with multiple addresses. Households with multiple physical addresses (e.g., snowbirds or households with weekend cabins) will have multiple chances of being selected from the ABS frame. To avoid the overrepresentation of such households that would otherwise result, a screening question will be used to enforce a “primary residence” rule.

Stage 4. Persons within Sample Addresses

The last stage of selection is done during the initial contact (household roster interview) of the sample address during the data collection phase. As with the NCVS, if the address is a residence and the occupants agree to participate, then an attempt is made to interview every person age 12 or older who lives at the resident address. The NCVS has procedures to determine who lives in the sample unit and a household roster is completed with names and other demographic information of all persons who live there (Attachment 9). These same procedures will be used across all conditions in the field test. Since the field test is a one-time survey, only those living at the sampled address at the time of enumeration will be included. If an age-eligible person leaves the household before an interview is completed, that person will be considered a nonrespondent.

Weighting and Estimation

Household, person, and victimization data from the NCVS sample are adjusted to provide annual estimates of crime experienced by the U.S. population age 12 or older. Following the creation of base weights, the nonresponse weighting adjustment then allocates the sampling weights of nonresponding households and persons to respondents with similar characteristics. A ratio adjustment reduces the variance of the estimate by correcting for differences between the distribution of the sample by age, sex, and race and the distribution of the population by these characteristics. This also reduces bias due to undercoverage of various portions of the population.

Base Weights. The NCVSIRTP field test base weight for each address is the inverse of the probability of selection for that address. In computing the probability of selection, any release of reserve sample will be accounted for.

Weighting Adjustments. If all eligible units in the sample responded to the survey and reported crimes only within the reference period, the sampling base weights would produce unbiased estimates with reasonably low variance. However, nonresponse and other nonsampling errors are expected in all sample surveys, and the following post-data-collection weighting adjustments are designed to minimize their impact on the estimates.

Drop Point Subsampling. Some units in the ABS sample are subsampled because the sampled address is associated with multiple residences (with no distinguishing feature); these are referred to as drop point addresses. As described in the preceding discussion of sample selection, units at drop point addresses will be enumerated and sampled. The base weights of units at these drop point addresses will be adjusted as appropriate to account for the change in the probability of selection.

Household Nonresponse. Nonresponse is classified into two major types: item nonresponse and complete (or unit) nonresponse. Item nonresponse occurs when a cooperating household fails or refuses to provide some specific items of information. In the NCVSIRTP field test estimation process, the weights for all of the interviewed households are adjusted to account for occupied sample households for which no information was obtained due to unit nonresponse. To reduce bias, the household nonresponse adjustment is performed within cells that are formed using the following variables: noninterview cluster, CBSA/MSA status, urbanicity, race of the household reference person, and interview number groups for the address.

Within-household Nonresponse. A household is considered a response if at least one person within the household completes the NCVSIRTP screening interview. The interviewer then attempts to interview all persons age 12 and older within the household, but some persons within the household may be unavailable or refuse to participate in the survey. This within-household nonresponse adjustment allocates the weights of nonresponding persons to respondents. The starting weight for all persons within responding households is the same address-level base weight multiplied by any drop unit subsampling factor and the household nonresponse adjustment factor. If nonrespondents' crime victimizations are significantly different from respondents' crime victimizations, there could be nonresponse bias in the estimates. To reduce nonresponse bias, the within-household nonresponse adjustment cells are formed by characteristics that are correlated with response and crime victimization rates. This includes: top 22 states/region, age, sex, race and Hispanic origin, and relationship to household reference person (self/spouse or all others).

Ratio Adjustment

Distributions of the demographic characteristics derived from the NCVSIRTP field test sample will be somewhat different from the true distributions, even for such basic characteristics as age, sex, race and Hispanic origin. These particular population characteristics are closely correlated with victimization status and other characteristics estimated from the sample. Therefore, the variance of sample estimates based on these characteristics can be reduced when, by the use of appropriate weighting adjustments, the sample population distribution is brought as closely into agreement as possible with the known distribution of the entire population, with respect to these characteristics. This is accomplished by means of ratio adjustments. The NCVSIRTP field test ratio adjustment has three high-level steps: (1) person coverage, (2) person iterative raking, and (3) household coverage.

Series Victimitizations. When a respondent reports a series crime (i.e., high-frequency repeat victimizations that are similar in type but occur with such frequency that a victim is unable to recall each individual event or describe each even in detail), the interviewer will complete one CIR for all of the incidents with details collected on only the most recent incident. In order to count all instances of this incident, the victimization weight will be multiplied by the number of times (up to 10) the incident occurred. Including series victimizations in NCVS national rates results in large increases in the level of violent victimization; however, trends in violence are generally similar regardless of whether series victimizations are included.

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

Total Crime Estimates

The NCVSIRTP field test data will allow users to produce estimates of crime and crime rates. Point estimates of crime victimizations include all incidents reported by sample units within the domain and time period of interest, weighted appropriately. NCVSIRTP field test crime rate estimates are calculated as the number of victimizations per one thousand people.

Variance Estimates

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

Replication methods may be used to estimate sampling error variances of survey estimates (and related measures of precision). For the NCVSIRTP field test, replicates will be created by collapsing strata. Each certainty PSU will comprise its own variance stratum (with SSUs combined to form two variance units within each variance stratum); noncertainty strata will be combined (paired) to form variance strata, with each noncertainty PSU corresponding to a variance unit. The sampling base weights are multiplied by replicate factors to produce replicate base weights. Each set of replicate base weights is subjected to the same weighting adjustments described in the previous section to produce sets of final replicate weights for households, persons, series victimizations, and incidents. By applying the weighting adjustments to each replicate, the final replicate weights reflect the impact of the weighting adjustments on the variance.

2. Procedures for Collecting Information

The field test is expected to be in the field from October 2019 through March 2020 (Conditions 1 and 2) and from January 2020 through August 2020 (Condition 3).

Data collection

For each of the three experimental conditions, the household roster instrument will be used to complete a **household roster** with names and other demographic information of household members (Attachment 9). Respondents will be asked to report victimization experiences occurring in the 12 months preceding the month of interview. In Condition 3, the roster interview will include questions about preferred modes of contact for the web interview. It will also include 5 broad questions about victimization and perceptions of neighborhood safety. These questions are intended to engage the respondent to improve response to the roster interview. The responses will also be used to assess nonresponse bias in the web survey administered two months later.

For Conditions 1 and 2, the household respondent will then be asked for **permission to record** the remainder of the interview (included in the consent language, see Attachments 3a, 3b, and 3c) and, if he or she is the parent of one or more sampled youth in the household, will be asked for **consent or permission to ask the youth to participate in the interview and for recording the interview** (Attachments 3c and 11). As other household adults are asked to participate, they will also receive these two brief modules as appropriate. The audio recordings will be used to detect potential falsification and validate interviews, as well as to edit survey data based on responses to an open-ended prompt for a description of a reported victimization

The **victimization screener**, either the current or the redesigned screener, will be asked of all respondents age 12 years or older in the household. It ascertains whether the respondent has experienced a personal crime victimization during the prior 12 months and is therefore eligible to be administered the **crime incident report (CIR)**. In each household, one respondent is designated as the head of the household and is asked about all

household property crimes on behalf of the entire household. The victimization screener collects the basic information needed to determine whether the respondent experienced a crime victimization (rape or other sexual assault, robbery, aggravated or simple assault, personal larceny, burglary, motor vehicle theft, or other household theft). The victimization screener for the redesigned instrument performs the same function. It asks about the above crimes as well as vandalism, and also broadly classifies reported victimization incidents to guide skip patterns in the CIR. See Attachment 12 for condition 1 (current NCVS, interviewer-administered) instrument and Attachment 13 for conditions 2 (redesigned NCVS, interviewer-administered) and 3 (redesigned NCVS, self-administered) instrument.

When a respondent reports an eligible personal victimization, the CIR is administered to collect detailed information about the crime incident. The CIR is administered for each incident the respondent reports, up to a maximum of four in the field test. For respondents reporting more than four incidents that do not comprise a series crime, the most serious incidents will be selected for the CIR. For each victimization incident, the CIR collects information about the offender (e.g. sex, race, Hispanic origin, age, and victim-offender relationship), characteristics of the crime (including time and place of occurrence, use of weapons, nature of injury, and economic consequences), whether or not the crime was reported to police, and victim experiences with the criminal justice system. The redesigned CIR includes expanded sections on behaviors and tactics related to sexual violence, reactions by victims to attacks, interactions with the police, and use of victims' services.

Another difference between the current NCVS and the redesigned instruments is that before the redesigned victimization screener each respondent will be asked a series of "non-crime" questions. The field test will include two such series, one on neighborhood safety and one on perceptions of the police. To minimize burden, each sampled individual will be assigned one of the two sets.

Finally, respondents in all conditions will be asked a few **questions about the interview experience** (Attachment 14). Responses will support assessment of the perceived burden and privacy of the interview.

The first contact with a household in all conditions will be by personal visit; for Conditions 1 and 2, subsequent contacts may be by telephone. For Condition 3, all contacts needed to complete the household roster interview will be in person. Two months after the household roster interview, sampled adults and youth with parental consent will be invited to complete the web survey ("non-crime" questions, victimization screener, and CIRs as appropriate). These initial **invitations** may be by surface mail, e-mail, or text message, as preferences were indicated in the roster interview (Attachments 5a, 5b, and 15). The intent of the two-month delay for Condition 3 is to simulate how respondents may react at the second time-in-sample interviews, including whether they are willing to complete the survey without an interviewer present and the effect of self-administration on data

quality.

3. Methods to Maximize Response Rates

Contact Strategy

Westat will mail an **introductory letter** explaining the NCVSIRTP and “frequently asked questions” (FAQs) to the household before the interviewer's first visit (Attachments 4a, 4b, 10a, 10b, 10c, and 10d). When they go to a household, interviewers will carry badges identifying themselves as Westat employees. Potential respondents are assured that their answers will be held in confidence and are used for statistical purposes. For respondents who have questions about the NCVSIRTP or the field test, interviewers provide a **brochure** (Attachment 7), and can also reference the FAQs. If there is no one at home on the first visit, interviewers will leave a **“Sorry I Missed You” card** (Attachment 16).

Westat trains interviewers to obtain respondent cooperation and instructs them to make repeated attempts to contact respondents and complete all interviews. Response rates will be monitored weekly, and interviewers will be coached on obtaining cooperation by their supervisors as needed.

Household nonresponse occurs when an interviewer finds an eligible household but obtains no interviews. Person nonresponse occurs when an interview is obtained from at least one household member, but an interview is not obtained from one or more other eligible persons in that household. Maintaining a high response rate involves the interviewer's ability to enlist cooperation from all kinds of people and to contact households when people are most likely to be home. As part of their initial training, interviewers are exposed to ways in which they can persuade respondents to participate as well as strategies to use to avoid refusals. Communications during the field period reinforce this training and offer further tips on obtaining cooperation.

Once the household roster is completed, the interviewer will attempt to complete the person-level interview (victimization screener and CIR as applicable for Condition 1; non-crime items, victimization screener, CIR (as applicable), and respondent debriefing for Condition 2) with the household respondent, and then with any other available adults or youth. For Condition 3, the interviewer will obtain contact information (mailing address, telephone number, e-mail address) for all adults and for youth for whom consent has been obtained. During the household roster interview for condition 3, interviewers will provide all households with a nonmonetary incentive (i.e., a magnet) to serve as a reminder of the survey when they are contacted two months after the household roster interview.

If a household initially declines to participate, an interviewer will return to try again unless the initial refusal is hostile or abusive. Similarly, if an individual declines to participate after the roster has been completed in Conditions 1 and 2, an interviewer will try again unless the initial refusal is hostile or abusive.

About two months after the Condition 3 household roster interview, sampled adults and youth with parental consent will be sent a surface mail letter invitation to complete the person-level survey. An experiment will be embedded in Condition 3 to test the effects of a promised incentive on survey completion and data quality. Following the Dillman web-push methodology (2017) where cell numbers and/or e-mail addresses are available, they will also be sent the invitation by text message or e-mail with a link to the web questionnaire³. Initial non-responders will be sent up to two reminders by surface mail, e-mail, or text. At that point, interviewers will be assigned to prompt (to complete the survey via web) respondents by telephone where numbers are available, and then in person. All Condition 3 person interviews will be self-administered on the web. In some cases, the respondent may wind up using the interviewer's laptop.

Individuals refusing to provide contact information or actively declining to participate after receiving the initial invitation (that is, contacting Westat or BJS by telephone, surface mail, or e-mail) will not be recontacted. Those declining to participate after being contacted by an interviewer may be recontacted, again excluding hostile or abusive refusals.

Interviewer Training

Training for NCVS interviewers consists of classroom and on-the-job training (Attachment 16). Initial training for interviewers consists of 40 hours of pre-classroom self-study and a 1-day classroom training, post-classroom self-study, and on-the-job observation and training. Initial training includes topics such as protecting respondent confidentiality, gaining respondent cooperation, answering respondent questions, proper survey administration, use of systems to collect and transmit survey data, NCVS concepts and definitions, and completing simulated practice NCVS interviews. This training is reinforced by written communications as needed, ongoing feedback from observations of interviews by supervisors, and regular performance and data quality feedback reports.

Monitoring Interviewers

In addition to the above procedures used to ensure high participation rates, Westat implements additional performance measures for interviewers based on data quality standards. Westat uses several corporate tools for monitoring interviewers: Computer Audio Recorded Interviews (CARI); Efficiency Analysis through Geospatial Location Evaluation (EAGLE) tool; Paradata Discovery & Decision Dashboard (PD3). Both EAGLE and PD3 provide visualizations of data available to supervisors for review on a daily basis. For the field test, managers will monitor interviewers using each of these tools, as shown in Table 5.

³ Dillman, D.A. 2017. The promise and challenge of pushing respondents to the Web in mixed-mode surveys. *Survey Methodology*, Statistics Canada, <http://www.statcan.gc.ca/pub/12-001-x/2017001/article/14836-eng.pdf>

Table 5. Westat field management tools and quality criteria for monitoring interviewer performance

Method	Quality Criteria	Feedback cycle
CARI: All interviews will be recorded, and a sample selected for review	Falsification detection based on presence of two speakers and the match between coded responses and what is heard; compliance with standardized interviewing rules, evaluated as reading questions as worded without changing meaning	Within 48 hours of receiving transmitted recordings
Geospatial Location Evaluation	Falsification detection based on comparison of location of interview to sampled address, using geocode matching	Discrepancies reviewed within 24 hours with interviewer
Productivity monitoring via status codes	Cooperation rates at household and person level	Monitoring weekly production goals against actuals per condition; supervisor follow-up within 48 hours of each refusal
Review of interview timings	Comparison of person level interview to median, outliers identified; interviews below threshold for minimum person level interview length	Supervisors will follow up on outliers and short interviews within 24 hours of identification
Monitoring of interview start times	Per time zone, interviews that start before 8AM or after 9PM	Supervisors will follow up within 24 hours of identification

Nonresponse and Response Rates

Anticipated response rates are shown in Tables 4a and 4b (Section 1). Despite the measures described in the preceding section, Westat does not expect to achieve the level of response experienced in the NCVS. There are two primary concerns for relative bias due to nonresponse: differences in the potential for bias across the three experimental treatments in the field test, and differences between the field test and the current NCVS. Analysis of field test results will include nonresponse bias analyses to explore both of these possibilities and assess possible effects on victimization estimates. At the household level, these analyses will use information available from the sampling frame or secondary sources such as the ACS. At the person level, nonresponse analyses will also include demographic information captured in the household roster interview.

4. Final Testing of Procedures

The NCVSIRTP cooperative agreement has included substantial formative and testing activities, including:

- Literature review of methodological studies involving the NCVS questionnaire;
- Review of publications using NCVS data, to assess the utility of existing questions;
- Observation of live NCVS interviews;
- Formative studies with web panels to aid in the development of cues for the victimization screener;
- Consultation with stakeholders and with substantive and methodological experts as detailed in Part 1;
- Three rounds of cognitive testing of the revised victimization screener;
- Two rounds of cognitive testing of the CIR; and
- Two rounds of usability testing of the self-administered questionnaire.

To support final preparations for the field test a final round of cognitive interviews was conducted to gather feedback and reactions to the final draft of the interviewer-administered redesigned instrument. Cognitive probes were focused on items that had been revised or added following two rounds of usability testing. A total of 15 cognitive interviews were conducted with adult victims of crime between in June–July, 2019. Respondents were randomly assigned to either receive the interleaving version of the screener or the non-interleaving version; and were randomly assigned to either receive the non-crime police or the community measures items.

Evidence from this final round of cognitive testing indicated that overall the final version of the questions, including the revised or new items, performed well. The questions were well understood, easy for interviewers to administer, easy for respondents to understand and answer, and captured the intended information. For the non-crime community measures, and revised scale was tested that included a middle category (e.g., somewhat or moderately worried. Respondents had a difficult time with the two qualifiers on this category, therefore it was recommended to move forward with using “somewhat” alone. The changes to the unwanted sexual assault contact screener generally tested well, but one key recommendation was to address threatened actions that may not have been captured in the screener. From this recommendation, the item S_07A2 was edited to specify “try or threaten to do this.” Finally, two new questions were added to the unwanted sexual contact section of the CIR to address whether the respondent had told the offender no or to stop, and if they had, whether the offender stopped. In the cognitive testing, the respondents had a difficult time understanding whether the question was

meant to capture whether the offender stopped at any point or if they stopped right away. To address this, the qualifier (i.e., “When you said this”) was moved to the beginning of the question. Additionally, this question will only be asked of attempted unwanted sexual contacts. The remaining recommendations from this cognitive testing were to address interviewer instructions, typos, programming edits, and consistency edits throughout the instrument.

In addition, Westat will conduct a pilot test of the field procedures for Conditions 1 and 2 in July–August, 2019. Clearance for this activity has been requested under separate cover (OMB No. 1121-0339). The pilot test is primarily a test of data collection field procedures, the computer-assisted instrumentation, and case management systems, and not of the survey instrument content, which will be a focus of the larger field test. The pilot test will inform the field protocols (i.e., mailing, household enumeration, consent, and interview procedures) and training needs for the larger field test. Pilot test data will be used to assess contact and cooperation rates, identify potential problems with the questionnaires, such as instrument routing, and item nonresponse. Based on the pilot test, any potential changes that may be needed for data collection field procedures will be implemented prior to the administration of the field test.

5. Contacts for Statistical Aspects and Data Collection

The Victimization Statistics Unit at BJS takes responsibility for the overall design and management of the activities described in this submission, including developing study protocols, sampling procedures, questionnaires, and overseeing the conduct of the studies and analysis of the data by contractors. Devon Adams is the Acting BJS Victimization Statistics Unit Chief.

Westat is responsible for the collection of all data. Ms. Wendy Hicks is the NCVSIRTP Project Director who will manage and coordinate the NCVSIRTP field test. BJS and Westat staff contacts include:

<p>BJS Staff: <i>all staff located at-</i> 810 7th Street NW Washington, DC 20531</p>	<p>Westat Staff: <i>Headquarters at-</i> 1600 Research Blvd. Rockville, MD 20850</p>
<p>Devon Adams Acting Chief Victimization Statistics Unit 202-353-3328</p>	<p>David Cantor, Ph.D. Vice President and Associate Director NCVSIRTP Principal Investigator 301-294-2080</p>

Jennifer Truman, Ph.D. Statistician Victimization Statistics Unit 202-514-5083	Jill Montaquilla DeMatteis, Ph.D. Associate Director NCVSIRTP Statistical Lead 301-517-4046
	W. Sherman Edwards Senior Advisor NCVSIRTP Co-Principal Investigator 301-294-3993
	Wendy Hicks Associate Director NCVSIRTP Project Director 301-251-2299