Implications of Repeated Sampling in a Crime Survey

J. Michael Brick[1], Grace Kena[2] , Pam Broene[1],
[1]Westat, 1600 Research Blvd. Rockville, MD 20850; [2]Bureau of Justice Statistics, 810 Seventh St., NW Washington, DC 20531

Abstract

One reason for conducting a longitudinal or panel survey rather than a repeated cross-sectional survey is to reduce the sampling error for estimates of change over time. Repeatedly sampling the same addresses has both operational and statistical effects. In particular, repeated sampling can affect response rates and reported outcomes (conditioning effects). The Bureau of Justice Statistics (BJS) uses a panel design in its National Crime Victimization Survey, the nation's primary source of data on crime reported and not reported to police. As one way of meeting growing stakeholder demand for subnational data, BJS explored the feasibility of a lower-cost mail survey that local jurisdictions could field to produce reliable information on victimizations and community attitudes about safety and local police agencies. The standardized format of the instruments permits comparisons to other areas and over time. The prototype of this National Crime Victimization Survey – Companion Survey (NCVS-CS) was tested in 2015 and again in 2016 in the 40 largest metropolitan areas in the U.S. In each year, a mail survey was sent to a probability sample of addresses. This research examines the advantages and disadvantages of overlapping the sample for estimating crime victimization and other related characteristics in this new survey. One-quarter of the addresses sampled in 2015, selected randomly, were retained in the sample for 2016. We examine the effects of retaining the same addresses on response rates, estimates of level and change for touched-by-crime rates and community attitudes, and the year-to-year correlations for these estimates. We give recommendations for designing the NCVS-CS and similar surveys, and for computing both level and change estimates.

Key words: Correlation over time, Optimal sampling rates, Response Rates, Conditioning Effects

1. Introduction

An important goal for many longitudinal or panel surveys is to reduce the sampling error for estimates of change over time. Longitudinal surveys sample the same units over time and, when the correlation of the estimates over time is large and positive as is often the case, the sampling error of change estimates are often less than they are for estimates from independent cross-sectional surveys. Longitudinal surveys may also have other operational and statistical effects. In particular, repeated sampling can affect response rates and response errors including conditioning effects from previous waves and bounding effects on estimates involving recall. This research examines the advantages and disadvantages of overlapping the sample for estimating crime victimization and other related characteristics for a newly developed mail sample survey.

As part of a multi-pronged strategy for producing crime data at the subnational level, the Bureau of Justice Statistics (BJS) explored the feasibility of a lower-cost mail survey that local jurisdictions could field to produce reliable information on victimizations and community attitudes about safety and local police agencies. The standardized format of the instruments permits comparisons to other areas and over time. BJS sponsors the National Crime Victimization Survey (NCVS) to produce national estimates of victimizations; the NCVS is the nation's primary source of data on crime reported and not reported to police. The data collection for this rotating panel survey is done each year by the Census Bureau using a face-to-face interview in the first wave of data collection and then retaining sampled addresses for 6 more interviews that are conducted every 6 months, largely using telephone. The NCVS is too complex and expensive to be conducted at the local level, so other, lower-cost alternatives are being considered.

One alternative design is a mail survey that could be conducted by state or local jurisdictions or their vendors. The BJS and Westat have been researching this approach and developed a prototype of this survey called the National Crime Victimization Survey – Companion Survey (NCVS-CS). The NCVS-CS was tested in 2015 and again in 2016 in the 40 largest metropolitan areas in the U.S. In each year, a mail survey was sent to a probability sample of more than 200,000 addresses. One-quarter of the addresses sampled in 2015, selected randomly, were retained in the

sample for 2016. This overlap experiment allowed investigation of the advantages and disadvantages of including the same addresses in the sample in both years.

The rest of this paper examines the effects of retaining the same addresses on response rates, estimates of level and change for touched-by-crime rates (a measure of crime victimization) and community attitudes, and the year-to-year correlations for these estimates. The paper concludes with recommendations on retaining sample addresses over time for the NCVS-CS.

2. Design

The primary objective of the NCVS-CS Field Test was to assess whether a mail survey could produce victimization estimates that correlate with those from the core NCVS (the annual survey fielded by the Census Bureau) and the Uniform Crime Report (UCR). The Field Test was conducted in two waves, a year apart. The Year 1 (2015) Field Test was a cross-sectional addressed-based sample (ABS) in the 40 largest core-based statistical areas (CBSAs) in the U.S., and was conducted between September and December 2015. The Year 2 sample was conducted between September and January 2017 and included both a new, independent cross-sectional ABS sample and an overlap sample of addresses from Year 1. A constant sample size was selected in each of the 40 CBSAs, except larger samples were selected in a few CBSAs to assess the feasibility of within-CBSA sampling and estimation. The oversampled CBSAs were stratified by police jurisdictional boundaries.

Both years included experiments to guide the development of future applications. One very important experiment that was conducted in both years compared two different ways of constructing the mail questionnaire. An **incident-level survey (ILS)** retained the approach used in the core NCVS with victimization probes followed by questions about reported incidents. The response structure is to ask about incidents and then link those incidents to the adults in the household who experienced them. The other questionnaire is called a **person-level survey (PLS)** because it asks about each adult and their victimizations, changing the focus from the incident to the person. The PLS begins by asking about property crime at the household level and then moves to asking about each adult and any violent crime they may have experienced. A second questionnaire experiment conducted in Year 1 varied the placement of questions on perceptions of community safety; this is called the version experiment. One version, Form A, placed these items at the beginning of the instrument, while the other version, Form B, placed them closer to the end after the questions on victimization.

The questionnaire design experiments were important because the alternative approaches could exhibit differential effects on unit response rates, completeness of reporting in households with more than one adult, item missing rates, and key outcome estimates including the number of adults who experience victimizations. In Year 2, the instrument experiment (ILS/PLS) was continued but the version experiment was discontinued because the analysis of the Year 1 data showed that Form A had distinct advantages over Form B, especially for the ILS. Williams et al. (in press) give details on the Year 1 questionnaire experiments.

The Year 2 also included a variety of new experiments. One of these new experiments was a test of retaining sample addresses for the second year, called the overlap experiment, and this is the main focus of this report. Since the version experiment was not fielded in Year 2, all of the addresses sampled for Form B in Year 1 were not eligible for the overlap. Only sample addresses with Form A in Year 1 were eligible for the overlap sample. A sample of 50 percent of the Form A Year 1 addresses were randomly selected and retained for Year 2. The rest of the sample was a new independent cross-sectional sample. If a sample address was eligible and retained from Year 1, it was administered the same instrument as in Year 1 (ILS/PLS) to avoid any respondent confusion. All of the sample addresses retained from Year 1 were included in Year 2, including refusals and vacants. A handful of cases with 'strong' refusals in Year 1 were not sent the Year 2 mailing and were classified as Year 2 refusals.

The other Year 2 experiments were mainly related to exploring the effects of using different data collection methods such as incentives and mailing conditions. Williams et al. (2018) discuss some of these issues with respect to the effectiveness of the mailing methods.

The Year 1 questionnaire experiments were randomized in a 2x2 randomized block design corresponding to the two levels of the instrument (ILS/PLS) and the version (A/B). The randomized units were defined by area (CBSA and

stratum within the oversampled CBSAs). Each area had an equal sample size of the treatments (PLS-A, PLS-B, ILS-A, ILS-B). The overlap experiment was also randomized within the same units. The ILS/PLS treatment condition for the overlap sampled cases was retained from Year 1 so there is no ability to examine an interaction for these two experiments for the reasons noted above. The other operational experiments also used a randomized block design, but are not discussed here. The sample size for the non-overlap Year 2 sample was 146,225 and the overlap sample size was 48,655.

Before giving the results with respect to the overlap experiment, it is valuable to review the performance of the NCVS-CS for eliciting responses from the sampled households. Across all experimental conditions and the 40 CBSAs, the Year 1 sample included almost 230,000 addresses with just under 94,000 households returning a survey. The overall response rate is 47.1 percent using the American Association of Public Opinion Research (AAPOR) method RR3[1] (AAPOR 2016). This response rate is considerably higher than could be achieved using other low-cost data collection methods such as telephone or the Internet. This rate is especially encouraging for a topic that is potentially sensitive. Further, large metropolitan areas are typically associated with lower response rates than national samples. The Year 2 sample included 217,250 addresses with just over 71,000 households returning a survey. The overall response rate is 40.9 percent using the AAPOR method RR3. This response rate is lower than the Year 1 response rate, at least in part due to the experimental treatments and overlap sampling as described below. The Year 2 response rate is still much higher than the approximately 15 percent response rate obtained when a telephone data collection was attempted in the pilot phase of this project (Edwards, Brick and Lohr 2012).

3. Response Rate Findings

While the AAPOR RR3 response rate is the generally preferred approach, for reporting response rates by experimental conditions AAPOR RR1 will be used. RR1 is a minimum response rate because it does not allocate the cases that remain with unknown residential status (not returned sample cases could be households or vacants), while RR3 does an allocation. The Year 2 response rate for the new (independent) sample that did not retain any addresses from Year 1 was 37.3 percent (RR1) compared to 33.2 percent for the overlap sample cases. These are the average response rate over the 40 CBSAs. The 4.1 percentage point difference is statistically significant ($p<0.05$) using a paired t-test with the overlap and non-overlap response rates for each CBSA as the observed values. In terms of response rates, retaining addresses from one year to the next reduces response rates.

Another approach to the analysis is to use a mixed logistic regression model to test the significance of the effects of the experiments. Table 1 shows the outcomes of the model and the overlap had a statistically significant effect with a lower response rate for the addresses retained from Year 1. The interaction of the other experimental conditions with the overlap were examined using this approach and no interaction was either statistically significant or substantively of interest. Thus, retaining sample addresses had the effect of reducing response rates independent of the other experiments in the study.

Table 1. Year 2 estimated coefficients of mixed logistic regression models predicting probability of responding, by experimental treatments and sampling frame variables

Effect	Estimate	S.E.	t value
Intercept	-0.748	0.044	-16.9
ILS	-0.003	0.010	-0.4
Incentive – $0	-0.359	0.012	-30.5
Incentive – $1	-0.138	0.012	-11.9
FedEx – None	-0.223	0.010	-23.2
Overlap – None	-0.190	0.011	-16.9

Note: The table shows the reference cell for the coefficients.

[1] Approximately 88% of the unknown eligibility addresses are likely to be eligible based on typical vacant rates in household surveys. If the AAPOR RR3 definition with this eligibility rate is used, then the response rate would be 47.9%; the minimum rate (AAPOR RR1) is 43.9%.

A key question in assessing the potential for nonresponse bias is whether the lower response rate by the overlap condition is related to any of the key outcome variables. For the NCVS-CS, the research question is whether responding addresses in Year 1 that are retained in the sample respond at a higher or lower rate in Year 2 depending on whether they reported a victimization in Year 1. Because being victimized is a relatively rare event, the number of addresses in the Year 1 sample that reported being victimized is small and only about 25 percent of these were randomly assigned to be in the overlap. Thus, the power to detect differences in Year 2 response rates by the Year 1 victimization status is limited. Nevertheless, the results are informative. For this analysis we use the touched by crime (TBC) outcome, that is '1' if any crime is reported in the household. For households reporting that at least one adult in the household was TBC for a violent crime in Year 1, the average response rate in Year 2 was about 14 percentage points lower than those households who did not report a violent crime. This difference, although large and of substantial importance, is not statistically significant. At the CBSA and strata level, about 80 percent had a lower nominal response rate for those touched by a violent crime compared to those not touched by a violent crime. Using the touched by property crime statistic, the pattern was similar; the average Year 2 response rate for those overlap household reporting being touched by a property crime in Year 1 was just under 7 percent lower than those who did not report a property crime. Again, this difference was not statistically significant, but the pattern across CBSA and strata exhibited a consistently lower response rate for those addresses reporting property crime in Year 1 (roughly 80% of the CBSAs had lower response rates for those who reported a property crime in Year 1).

Thus, retaining the units from Year 1 to Year 2 depressed response rates in Year 2 and there was some indication, although not statistically significant, that the households reporting victimizations in Year 1 were less likely to respond in Year 2 than those households that did not report a crime in Year 1.

4. Overlap Effect on Victimization and Other Outcomes

In this section we extend the analysis to give a more detailed analysis of the effect of the overlap on the key estimates. For this purpose, the touched-by-crime (TBC) statistics were computed by the instruments and other experimental treatments. For this analysis, three TBC variables are presented: TBC-property crime, TBC-violent crime, and TBC-serious violent crime. These can be computed at the household and adult level. For this analysis, the property crime is at the household level and the violent and serious violent crime are at the adult level. These same statistics are used for both Year 1 and Year 2.

For each CBSA, the final raked weights were used to calculate TBC rates for each CBSA, separately for each instrument (ILS and PLS). The instrument did have an effect on the rates so all the analysis given below is done separately by the instrument. As was done in the section on response rates, the 40 CBSAs were the units at which the rates were computed and then the rates were summarized by averaging over the CBSAs. The main analytic method used was a paired t-test where the TBC rate for each CBSA was computed (separately for ILS and PLS) for the overlap and non-overlap samples. The overall significance was computed based on the 40 differences.

Table 2 shows the differences for the ILS and Table 3 shows the differences for the PLS. For both instruments, the average percentage of households reporting a property crime was about 2 to 3 percentage points lower for the overlap sample respondents compared to the non-overlap respondents. For violent and serious violent TBC rates at the person level, the overlap estimates were again significantly lower than the non-overlap estimates for both the ILS and PLS.

Table 2. Year 2 paired t-test comparison of TBC rates for retained and new addresses, for ILS

TBC variable	Average overlap percent	Average non-overlap percent	Difference	p-value for difference = 0
Property	9.6	12.1	-2.5	<0.01
Violent	1.3	1.8	-0.5	<0.01
Serious violent	1.1	1.5	-0.4	<0.01

Table 3. Year 2 paired t-test comparison of TBC rates for retained and new addresses, for PLS

TBC variable	Average overlap percent	Average non-overlap percent	Difference	p-value for difference = 0
Property	13.3	16.1	-2.8	<0.01
Violent	2.0	2.5	-0.5	0.05
Serious violent	1.2	1.6	-0.4	0.04

In addition to the victimization items, some items on community safety and attitudes to policing (CPQ) were included in both Year 1 and Year 2. Three of these items were dichotomized and are analyzed below. The three variables are:

- SAFE - On the whole, how much of the time is the community where you live safe? (always safe or mostly safe versus sometimes safe, rarely safe, or never safe).
- FEAR - How often does fear of crime prevent you from doing things you would like to do? (rarely or never versus very often or somewhat often).
- RATING - Taking everything into account, how would you rate the job the local police are doing? (very good job or somewhat good job versus neither good nor bad job, somewhat bad job, or very bad job).

Tables 4 and 5 show the estimates for the two years for these three items. Although the differences are not very large, the respondents in the overlap did report feeling safer in their community and having a higher rating of the police. These responses are consistent with having fewer respondents who reported a victimization in the overlap. People who have not been victimized tend to rate police higher than those who have been victimized.

Table 4. Year 2 paired t-test comparison of CPQ estimated percent for retained and new addresses, for ILS

CPQ variable	Average overlap percent	Average non-overlap percent	Difference	p-value for difference = 0
SAFE	85.5	84.4	1.1	0.02
FEAR	74.4	73.2	1.2	0.09
RATING	83.4	82.3	1.1	0.06

Table 5. Year 2 paired t-test comparison of CPQ estimated percent for retained and new addresses, for PLS

CPQ variable	Average overlap percent	Average non-overlap percent	Difference	p-value for difference = 0
SAFE	85.9	84.5	1.5	<0.01
FEAR	74.1	73.0	1.1	0.08
RATING	83.2	82.1	1.1	0.01

5. Estimates of Change

To estimate change from Year 1 to Year 2, the CBSA estimates for each year were computed and the differences in the estimates computed. This approach exploits the blocked design to provide an efficient estimate of change. Since the Year 2 survey dropped Form B, the differences were computed between the entire data set from Year 2 (all used Form A) and the statistics calculated using only Form A from Year 1. For each CBSA stratum, the change statistic is calculated as the percent TBC in Year 2 minus the percent TBC in Year 1 from Form A. Each percentage is calculated using the final weights.

Table 6 gives the average of the change statistics. The table shows different patterns for estimating change in TBC rates with the two instruments. The PLS shows a decrease, on average, in property crime between 2015 and 2016; the ILS indicates an increase. The PLS instrument added some questions on dates for property crimes in 2016 and this may have contributed to observed lower rates for property crimes. Both instruments indicated that there was an increase in violent crime between 2015 and 2016, but the increase was higher for PLS than for ILS, although the difference was only statistically significant at the 0.05 level for violent crimes.

Table 6. Average change from Year 1 to Year 2, for PLS and ILS, for victimization rates.

TBC variable	Average change, PLS	Average change, ILS	PLS change - ILS change	p-value for difference = 0
Property	-2.0	0.5	-2.5	<0.01
Violent	0.7	0.3	0.4	0.03
Serious Violent	0.2	0.2	0.0	0.94

A similar analysis was done for the CPQ items discussed earlier and the PLS and ILS change estimates were not statistically significant. A key result is that the one-year change estimate for both victimization and attitudes is small, as might be expected. The sample sizes in each of the CBSAs is relatively large, but is not large enough to detect changes in characteristics, especially for rare characteristics that are based on only a few households reporting a TBC per CBSA.

Table 7 displays the correlations between the Year-1 TBC rates and the Year-2 TBC rates. These are calculated as the correlations between the 40 CBSA-level statistics. Each of the CBSA-level statistics was calculated using the weights. The correlations are between the TBC rate in the CBSA from Year 1 and the TBC rate from Year 2, calculated using the PLS respondents and then using the ILS respondents. The correlations between the Year-1 TBC and the Year-2 TBC for property crime are high for both PLS and ILS. However, the correlations between the Year-1 TBC and Year-2 TBC for violent crime and serious violent crime are low and are not statistically significantly different from zero. Violent crime is a rare event, and many of the CBSAs had only a handful of violent crimes reported on the questionnaires so a difference of one or two violent victimizations in a CBSA would result in a different rate and could reduce the correlation.

Table 7. Correlation among Year-1 and Year-2 TBC rates, by instrument.

TBC variable	PLS correlation	p-value	ILS correlation	p-value
Property	0.89	<.01	0.81	<.01
Violent	-0.00	0.9	0.17	0.30
Serious Violent	0.02	0.93	0.14	0.41

Table 8 shows analogous correlations for the 3 CPQ items. For these, the Year-1 and Year-2 statistics from both PLS and ILS show high correlations.

Table 8. Correlations among Year-1 and Year-2 CPQ estimates, by instrument.

CPQ variable	PLS correlation	p-value	ILS correlation	p-value
SAFE	0.74	<.01	0.73	<.01
FEAR	0.88	<.01	0.81	<.01
RATING	0.54	<.01	0.45	<.01

Having high positive correlations is one essential ingredient required to achieve efficiency in estimating change over time in a longitudinal design. For both property crime and for the three CPQ items, a high correlation is observed. An important point to note is that the correlations studied here are at the CBSA level rather than at the individual household level so these correlations only suggest that the efficiency gains for estimating change might be gained in this situation. The next section takes a more direct analytic approach to study the gains due to retaining the sample addresses.

6. Gains in Efficiency

The method used to directly estimate the gain in efficiency for estimating change compares the estimated variance of the change using the overlap sample to the variance that would have been expected if the addresses in the two years from the overlap sample were independent. Let $\hat{p}_1$ be an estimate from Year 1 from the overlap sample households, and let $\hat{p}_2$ be a corresponding estimate from Year 2. The variance of the change estimate is $V_O(\hat{p}_2 - \hat{p}_1) = V(\hat{p}_2) + V(\hat{p}_1) - 2\,Cov(\hat{p}_1, \hat{p}_2)$. If the sets of households in Year 1 and Year 2 were independent, then the variance of the change estimate would be $V_I(\hat{p}_2 - \hat{p}_1) = V(\hat{p}_2) + V(\hat{p}_1)$. If the ratio of V_O/ V_I is close to one, then the estimated covariance between the Year-1 and Year-2 estimates is close to zero and there is no precision benefit from retaining the same addresses for Year 2. If the ratio is less than one, then there is evidence that the overlap sample has smaller variance than two independent samples.

Table 9 shows a summary of the distribution of the ratio of V_O/ V_I estimated for the three TBC statistics; Table 10 gives the corresponding estimates for the CPQ items. In both tables, the means and medians of the variance ratios (measuring the center of the distribution) are both close to one. This finding suggests that for a "typical" CBSA, the variances were essentially the same as would have been expected with independent sample in the two years. These tables show no evidence of a reduction in variance for change estimates from retaining addresses in the sample.

These results may be somewhat surprising, especially for the CPQ estimates that have such high CBSA correlations as shown in the previous section. There are two main explanations for these results. One explanation is that the expected variance for the independent sample is based on a larger sample because the response rate for the new addresses is considerably larger (14%) as discussed earlier. This increased sample size more than compensates for any positive correlation. A second explanation is that the correlations at the household level are probably less than the estimates of the correlations shown in the previous section that are at the CBSA level. The ratio approach used here avoids having to estimate the correlations at the individual level.

Table 9. Summary of distribution of the ratio of V_O/ V_I for the TBC rates, by instrument.

TBC variable	25th Percentile	Median	75th Percentile	Mean
Property	0.91	0.99	1.02	0.97
Violent	0.96	0.99	1.04	1.00
Serious Violent	0.96	1.00	1.02	0.99

Table 10. Summary of distribution of the ratio of V_O/ V_I for the CPQ estimates, by instrument.

CPQ variable	25th Percentile	Median	75th Percentile	Mean
SAFE	0.97	1.04	1.10	1.02
FEAR	0.92	0.99	1.07	0.99
RATING	0.97	1.01	1.07	1.02

7. Conclusion

The NCVS-CS experimented with retaining a sample of addresses from the 2015 (Year 1) sample and asking the people at that address to respond to the same survey (with slight changes) the next year. The effects of the overlap were substantial and important. Retaining sample addresses for the next year:

- Reduced the response rate by about 4 percentage points compared to the new sample addresses;
- Appears to have affected response rates differentially, with those reporting a victimization in Year 1 being less likely to respond in Year 2 than those who did not;
- Produced lower estimates of victimization and higher estimates of feeling safe and police ratings in Year 2 than the new, independent sample; and
- Was no more efficient than the independent sample for estimating change from year-to-year.

Since one of the main goals of retaining sample across data collections is to reduce the variance of estimates of change, the experiment showed that this was not accomplished for the NCVS-CS. In some surveys, the cost of data collection is lower in subsequent waves of data collection because lower-cost modes can be used. This does not apply to the NCVS-CS as currently designed. The lower response rate and potential nonresponse bias due to differential Year 2 based on Year 1 victimization also raise concerns about using an overlap sample. The recommendation is that the NCVS-CS should not retain sample addresses over time.

These findings may have some relevance to other longitudinal surveys, but the circumstances of each survey need to be carefully considered. For example, the core NCVS is a rotating panel with 6 follow-up waves every 6 months after the first interview. In this survey, the follow-up waves are generally conducted by telephone which is considerably less expensive than the face-to-face first interview so cost is an important consideration. Another difference is that the key estimates from the core NCVS are incident rates rather than TBC rates. Incident rates are ratios where the numerator is the number of victimization incidents rather than the binary (0,1) numerators used in TBC rates. This difference could have an important effect on the correlations and the efficiency of retaining sample addresses. This research did not investigate the effect for these types of estimates.

References

.

The American Association for Public Opinion Research. 2016. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 9th edition. AAPOR.

Edwards, W. S., Brick, J.M., and Lohr, S. (2012).

Williams, D., Brick, J. M., Edwards, W. S., and Giambo, P. (in press). Questionnaire Design Issues in Mail Surveys of All Adults in a Household. In P. Beatty, D. Collins, L. Kaye, J.L. Padilla, G. Willis, and A. Wilmot (Eds.), Advances in Questionnaire Design, Development, Evaluation, and Testing. Hoboken, NJ: Wiley

Williams, D., Brick, J. M., Edwards, W. S., Giambo, P., and Kena, G. (2018). Cost Effective Mail Survey Design. Paper presented at the Federal Committee on Statistical Methodology in Washington, D.C.