Examination of Data Collection Methods for the National Crime Victimization Survey

Lisa Carley-Baxter[1], Emilia Peytcheva[2], Vincent Iannacchione[3], and Rachel Caspar[4]

[1, 2, 3, 4] RTI International

[1]lcbaxter@rti.org, [2]epeytcheva@rti.org, [3]vince@rti.org, [4]caspar@rti.org

Introduction

The National Crime Victimization Survey (NCVS) has become the Nation's primary source of information on the frequency, characteristics, and consequences of criminal victimization in the United States. Each year data are obtained from a nationally representative sample of approximately 38,600 housing units comprising nearly 67,000 persons. As the cost of collecting survey data through a face-to-face mode continues to increase, the NCVS has faced the challenge of maintaining adequate sample sizes to support analysis needs within the funding currently available.

The primary purpose of the current research is to identify new survey methods that will lower the cost per case for completed NCVS interviews while minimizing the impact on response rates and standard errors of key estimates. We will investigate three major avenues, as well as associated issues, for potentially reducing the costs for completing the NCVS: mixed mode data collection, reducing the number of persons in the household to interview, and using address-based samples to collect data in different modes.

Mixed Mode Experiment

We have designed an experiment to evaluate the impact of mixed-mode (in person, telephone, and web) data collection coupled with incentives on survey costs, response rates, and survey estimates, including an examination of potential mode bias. Table 1 details the experimental conditions.

The first experimental condition in Wave 1 aims to implement a protocol that closely resembles the current data collection. In this condition, labeled as condition 1 in Table 1, Computer-Assisted Personal Interviews (CAPI) are attempted with all sample addresses. Household interviews must be completed via CAPI, however for individual interviews household members could be interviewed by phone, by the field interviewers (as opposed to centralized CATI). This control condition is needed to ensure comparability between the baseline effort and the experimental conditions. Using the most current NCVS data would not provide comparable data as multiple survey factors impact the data collection. For example, response rates can be affected by the geographic area of the experiment, and similarly the interviewers, the recruitment procedures, how call outcomes are coded and response rates computed, and other differences between survey organizations and sample designs can also impact data outcomes.

Table 1. Experimental Design for NCVS.

		Wave 1		Wave 2	
Condition	Type of Contact	Household Respondent	Individual Household Members	Household Respondent	Individual Household Members
1	Initial Contact	CAPI	CAPI	Mail/Web and Inbound	Mail/Web and Inbound
	Follow-up	none	CATI	CATI	CATI
2	Initial Contact	CAPI	Mail/Web	Mail/Web and Inbound	Mail/Web and Inbound
	Follow-up	CATI	CATI	CATI	CATI
3	Initial Contact	Mail/Web	Mail/Web	Mail/Web inbound	Mail/Web inbound
	Follow-up	CATI/CAPI	CATI/CAPI	CATI	CATI
4	Initial Contact	Inbound and Outboud CATI	Inbound and Outboud CATI	Mail/Web inbound	Mail/Web inbound
	Follow-up	CAPI/CATI (if appt)	CAPI/CATI (if appt)	CATI	CATI

The same procedures are followed in the second condition in Wave 1, attempting CAPI interviews for all sample addresses, but only one additional household member is selected. This will permit the estimation of response rates to the household interview if sample members are informed that only one other household member will be interviewed. It will also allow the estimation of changes to the household member individual interview response rates, although confounded by any differences in cooperation to the household interview.

The third condition in Wave 1 aims to be the most cost efficient per completed interview. Sample addresses are mailed a letter and asked to go to a web address and complete the survey. A mail questionnaire may be enclosed, or mailed separately to nonresponding households. After the roster information is received, the same procedure will be followed for the other household members as a means of completing individual interviews. Remaining nonrespondents will be followed up by phone when a telephone number is available or provided by a household member, or in-person if a telephone number is not available. If the field period permits, in-person follow-up will be made to sample members in the telephone follow-up as well, as a means of increasing response rate.

Depending on the use of follow-up modes, this option may yield the lowest cost per completed case. However, this can come at the cost of substantially lower initial response rates, and possibly overall response rates. Further, while telephone and in-person interviewing can be used to increase cooperation, some sample members may no longer be willing to participate and the cost benefit may be lost through the extensive need to rely on these follow-up modes. There is a potential benefit if subsequent waves of data collection are also by self-administration, as the bounding baseline (Wave 1) interview will be in the same mode for many respondents.

However, data from the baseline interview are not directly used for producing population estimates and any reduction in differential measurement error can be overwhelmed by increases in coverage and nonresponse error, through factors such as increased error in household rosters, literacy, computer literacy, and lessened ability to adjust for attrition through greater nonresponse in the baseline interview.

Even more so than in the second condition, the creation of mail questionnaires is problematic – the household roster (control card) and household interview will both need to be designed for mail format. This design will also require a longer field period, as the household roster and household interview need to be received and processed before the mail out of individual household member questionnaires. The self-administered formats, mail and web, pose an additional concern at this initial contact as consent needs to be obtained in order to send questionnaires to or interview minors in the household, yet there is no control over who fills out the form and provides the consent.

An incentive will be highly desirable to motivate sample members to complete the self-administered survey instruments, yet this also poses additional threat to who fills out the questionnaires and consent form (i.e., minors, individuals who are not members of the household, and individuals who are not responsible for the child given consent).

The fourth condition in Wave 1 tests the use of only telephone and in person, subjecting as many cases as possible to telephone. Letters will be sent to all sample addresses asking them to call a toll-free number, to complete the roster and household interview. An incentive will be critical to the success of this design to motivate sample members to initiate the call, and would be offered if possible. After about three weeks, calls will be made to all phone numbers that can be matched to nonrespondent sample addresses (outbound CATI), a rate that will likely approach 50%. Even a greater proportion of the sample is expected to be attempted by telephone among the remaining household members, as household respondents would have provided telephone numbers. Finally, remaining nonrespondents will be attempted in-person for CAPI. As much of the CAPI cost is associated with travel and particularly with travel that results in noncontact, household members who are at home at the time of a CAPI interview will be asked to do the survey at that time in CAPI. This condition is particularly useful in the event that the use of self administration is found to be a poorly performing design. Additional efficiency can be achieved through the use of centralized CATI, particularly when implemented on a large scale, and although this efficiency may not hold for all surveys and survey organizations it is expected to occur under the current protocol.

One of the strengths of this option is that it does not rely on self administration and in the event that the designs involving web and mail modes fail to meet the study objectives, this design can still achieve cost savings by placing substantially greater emphasis on the telephone data collection than the current NCVS protocol. While the incentive will pose an expense, it should still lead to substantial savings from not having to make numerous call attempts to those cases, whether by phone or in-person. This design will require a complicated sample management system, but major complexities in some of the other design options are avoided - the instrument remains computerized and does not require changes to accommodate a static mode. Interviewer administration may also help build the rapport that would reduce attrition following the first

wave of data collection. A drawback may be that a large proportion of sample addresses still require in-person data collection.

We will also conduct a second wave of data collection with all cases in all four experimental conditions. We will attempt to complete household and individual level interviews via mail or web initially. Nonresponse cases will be followed up via CATI. We anticipate that conditions 1 and 2 will have the highest overall response rates because the in-person data collection in Wave 1 will serve a dual purpose of building rapport with the respondents, as well as collecting information on as close to the full sample as possible that can be used in adjustments for nonresponse and attrition. This also allows the collection of multiple types of contact information, such as e-mail addresses, and landline and cell phone numbers that can be used in subsequent waves of data collection and limits the types of spelling errors and missing data that could occur by telephone. While this design will likely involve the greatest cost in the first wave, it could yield savings in the subsequent waves of data collection that are used for producing NCVS estimates particularly when the first mode of contact in the subsequent waves is through mail or web, followed by telephone.

Optimal Number of Persons Interviewed per Household

Second, we will examine the optimal number of persons to interview in each household. In its current administration, the NCVS interviews all eligible persons in a household who are age 12 or older. The average number of persons interviewed is about 1.8 per household. A change in the number of persons interviewed could have far-reaching and unknown implications for the response rate, the survey estimates, the statistical efficiency of the design, and survey costs.

One benefit of the current NCVS design is that household respondents help to secure the participation of other household members by providing their contact information and times when they are likely to be available. Further, we suspect that if the household respondent's interview experience was positive, then he or she relays this fact to other household members, helping to secure their cooperation. Household respondents help to locate and motivate other household members to respond, a burden which would otherwise fall on interviewers. On the other hand, household respondents may also serve as gatekeepers, denying access to other household members if their interview experience was unpleasant.

If the NCVS were to switch to a design where only one person in the household is interviewed, the cost-free assistance provided by household respondents would be lost. Interviewer labor would be greater because a significantly larger number of households would need to be contacted and persuaded to participate, with more opportunities for refusals. Although we have no empirical evidence, we suspect that survey costs would increase and the response rate would decrease as a result of this change.

Survey costs and response rates cannot be the sole considerations in this decision, however. The present NCVS design is highly clustered within individual households, introducing non-negligible intracluster correlation (and higher variances) into the survey estimates. By eliminating the clustering effects, it may be possible reduce the sample size to save costs, or to maintain the same sample size but improve precision. Either of these outcomes would be desirable in terms of statistical efficiency.

One possible approach to estimating the change in victimization rates by switching to a single-respondent design involves using existing NCVS data. Using 2006 survey data, for example, we could randomly select one respondent per household, estimate victimization rates, and compare them to rates estimated using multiple-respondent households. For crimes that tend to affect all of the residents in a household (e.g., burglary), we would expect the one-person per HH design to have little loss of statistical efficiency because of intra-household correlation. For crimes that happen outside the home (e.g., assault) however the efficiency of the prevalence rates could be noticeably reduced.

An important caveat of this approach is that the existing NCVS data for multi-respondent households may be subject to certain biases for intra-familial crimes such as domestic violence. For example, a husband who has been interviewed may tell his wife not to report domestic violence. In this situation, the resulting under-reporting could be reduced with a single-respondent per household (assuming privacy is maintained in the interview setting).

We will supplement this research with an examination of the response rates achieved in other household surveys using single-respondent designs to observe if response rates are likely to increase or decrease if this design change is made. We will also calculate the costs associated with traveling to and gaining cooperation with a larger sample of households, and compare these costs to the current design. Taken together, this information will describe the costs and benefits of changing to a single-respondent design.

Feasibility of Address-Based Sampling to Support Different Data Collection Modes

Third, we will examine the feasibility of using address-based samples to collect data in different modes, in part by attempting to locate names and phone numbers using commercially available databases. We will use several batching vendors and increasingly targeted approaches to provide the highest rate of correct phone number matches. Initially, we will batch the sample through a phone number and name appending process. By appending name as well as phone number, we will develop an additional identifier to ensure that the end results are active phone lines. This batch process will match the sampled address with the occupants currently believed to be associated with that address.

The second tier of the batching process will involve using the names, phone numbers and addresses to conduct a comprehensive phone search. We will obtain up to three verified phone numbers for the provided address (including cell phone numbers) as well as the date of the most recent association of that number and person with the sampled address. Using this data, we can determine the most likely current occupants.

The third tier of the batching process will match the developed names, addresses and phone numbers with active telephone directories to further increase the likelihood that the developed number is active and associated with the sampled address. As well as ensuring the highest number of active phone numbers, this last step will further increase the number of phone numbers matched to the sample.

While we do not expect to retrieve telephone numbers for all the addresses on the frame, our experience suggest that between 50 and 60 percent will be successfully matched. As part of our experiment, we will monitor and report match rates, as well as the proportion of numbers that turn out to be incorrect or non-working.