Assessing Interactive Voice Response for the National Crime Victimization Survey

Final Report

Authors

David Cantor
Douglas Williams

Draft: August 29, 2013

Prepared for:
Bureau of Justice Statistics
810 Seventh Street, NW
Washington, DC 20531
Contract: 2008-BJ-CX-K066

Prepared by:
Westat
1600 Research Boulevard
Rockville, Maryland 20850-3129
(301) 251-1500

Westat®

Executive Summary

Introduction

The National Crime Victimization Survey (NCVS) relies exclusively on interviewer-administered modes of data collection. Since the redesign of the early 1990s, these modes have included in-person and telephone interviewing. The objective of this project is to examine the use of Interactive Voice Response (IVR) as a mode of data collection for the NCVS. There are several possible applications of IVR for the NCVS. As part of the NCVS core methodology, an IVR mode could be incorporated as part of a multi-mode design within the rotating panel design. For instance, after the initial in-person interview, respondents could be asked to call into an 800 number to complete the survey in subsequent contacts. By establishing rapport at the first interview, the NCVS could efficiently collect data from some portion of the sample without incurring the expense of an interviewer-driven methodology. A second possible use of IVR is for a supplemental survey to generate local area estimates. Previously, agencies have relied on mail or telephone surveys to conduct local area victimization surveys. IVR could provide a way to increase the efficiency, and possibly the quality, of these surveys.

The purpose of this project was to address three questions related to the use of an IVR on the NCVS: 1) What are the response rates with an IVR?, 2) Can the NCVS be adapted for an IVR interview and 3) What are the prevalence and victimization rates for an IVR?

Methods

These questions were address with an experimental design that varied the method of contact (telephone interviewer vs. mail), mode of response (speech vs. keypad), and methods to enhance the response rate (promised incentive vs. no incentive; insert vs. no-insert). Table A provides these variations.

Table A. Experimental Variations Tested on Project

		Response Entry Method			
		Speech		Keypad	
IVR Mail (Mail Contact)		$0	$10	$0	$10
Insert	Yes				
	No				
IVR Telephone (Phone Contact)					
CATI (Phone Interview)					

The IVR Mail condition started with sending a letter to the sampled address, requesting the adult with the next birthday to call a toll free number to complete the IVR. The IVR Telephone condition started with a phone call to the sampled household. Once a respondent was selected, the individual was switched to the IVR to complete the survey. The computer assisted telephone interview (CATI) condition was a telephone interview, administered by an interviewer, using the NCVS instrument.

The survey was carried out in the St. Louis and Houston metropolitan statistical areas. There were two sampling frames used to sample addresses: 1) all residential addresses based on United States Postal Service (USPS) administrative data, and 2) lists of addresses obtained from the St. Louis and Houston police departments of individuals who reported a crime within the prior 12 months. Sampled addresses were reverse-matched to find a telephone number. With a minor exception, only those addresses that had a match to a telephone number were retained in the sample.

Approximately 10,000 addresses were allocated to conduct the IVR Mail condition. Approximately 3,000 addresses were allocated to the telephone request to complete the IVR Telephone condition and 1,500 addresses were allocated to administer the CATI.

All sampled households were sent a $2 incentive as part of the initial letter sent to the household. The IVR Mail condition was administered with four different contacts: 1) an initial request to complete the survey, 2) a reminder postcard, 3) a second follow-up request and 4) a third follow-up request. The telephone contacts were all sent a pre-notification letter prior to calling and asking for participation on the survey.

Question 1: What are the response rates for the IVR?

The experiment examined two questions related to response rates. One was to compare response rates by different experimental conditions. The second was to test two methods to enhance the IVR response rate.

Comaprative response rates. The overall response rates for the three contact conditions were 23 percent for IVR Mail, 18 percent for IVR Telephone and 24 percent for the CATI. This experiment provided evidence that for NCVS-1 the IVR Mail methodology is equivalent to CATI. The IVR Telephone methodology had lower response rates. The keypad respondents had higher response rates than the speech respondents. With respect to respondent characteristics, there were very large differences between all of the experimental groups and the general population. When comparing to ACS data, all of the groups underrepresented young people, those with a lower education and Hispanics.

The IVR response rates drop once considering completion of NCVS-2. Overall, the IVR-Mail and IVR-Telephone rates for the ABS sample drop 2 – 3 percentage points. While this is a relatively small drop in the overall rate, this hides the fact that a significant number of individuals did not complete all of the NCVS-2 forms that were expected. Overall, approximately 30 percent of the respondents did not fill out all of the required incident forms. This compares to a 100 percent completion rate for the CATI.

Enhancements to the response rate. The results of the experimental manipulations found significant effects of a promised incentive of $10 to encourage use for the IVR Mail condition. When paired with the insert, the response rate went up by approximately 10 percentage points (32 percent compared to 23 percent). The promised incentive did not work as well for the IVR Telephone group. The incentive improved data quality. Respondents receiving the incentive were more likely to fill out all detailed incident forms. They also had less missing data on income. However, there was no evidence that the enhanced response rate significantly improved representation of demographic groups.

Implications. From a response rate perspective of completing NCVS-1, the IVR Mail condition is at least equivalent to a CATI interview. If used in conjunction with a keypad entry mode, an incentive and an insert, it performs significantly better. Completing the NCVS-2 is more problematic. A significant number of respondents did complete multiple forms. However, there was a significant drop-off in response, with 30 percent of respondents not completing all of the forms required of

them. These results suggest that the IVR Mail procedure is promising as either a follow-up to NCVS respondents (e.g., at T2 – T7) or as a way to screen respondents for a local survey.

Question 2: Can the NCVS be adapted for an IVR interview mode?

This question examined several aspects related to the IVR across NCVS-1 and NCVS-2. The first was to examine the overall usability of the IVR. The second was to examine how well the IVR completed tasks normally done by the interviewer for NCVS-1. The third aspect was to examine how well these tasks were completed for NCVS-2.

Usability. The average time to complete the NCVS-1 screener ranged from 7 to 10 minutes, depending on whether a victimization was reported. Once a victimization was reported, the time to complete goes up to between 16 and 24 minutes, depending on the number of NCVS-2s filled out. The speech respondents had the highest error rates, with the driving problem being the system failing to recognize their responses. On NCVS-1, 76% of the speech respondents had at least one question where the system did not recognize their response. There were also several problems with the design of selected IVR questions, including income and selected 'mark all that apply' items. Those who had at least one victimization were more likely to report being less satisfied with the questionnaire, confused, frustrated and that the survey was too long. Finally, there were instances where the IVR system did not correctly code the data.

Adapting the NCVS-1. Adapting the IVR for NCVS-1 was reasonably successful at collecting reports of victimizations. The procedures in place were successful at detecting duplicate incidents. Similarly, respondents did take advantage of the verification procedure at the beginning of the incident form to correct reports on the screener.

With respect to actually reporting victimizations, there was a small, but significant difference between the two modes of contact, with the IVR Mail having a higher proportion reporting victimizations than the IVR Telephone. This result holds up after controlling for differences in household and demographic characteristics. When compared to the CATI, the IVR had a nominally higher proportion of persons reporting a victimization, but the differences were not statistically significant.

The number of victimizations reported for the IVR was high. For the ABS sample, 40 percent of those reporting a victimization had at least two incidents. The number reported on the IVR was

significantly higher than the CATI. While there was a tendency for IVR respondents to reduce these numbers at the verification at the beginning of the NCVS-2, this only affected about 10 percent of the respondents reporting a victimization.

Adapting the NCVS-2. Adapting the NCVS-2 for IVR poses different challenges than NCVS-1. As adapted for this project, the IVR successfully carried out many of the required functions. To assess the quality of the NCVS-2 data, all summaries contained at the end of the incident form were reviewed and compared to the initial TOC assignment. This review found that most respondents were able to successfully navigate the transition between NCVS-1 and NCVS-2. There were some respondents where this transition did not successfully occur for one of several reasons: 1) some respondents did not cleanly identify the incident that was targeted for the NCVS-2 report, 2) some respondents reported on events that were clearly outside the reference period, and 3) some respondents reported events that were not committed against an ineligible victim.

For purposes of classifying incidents into a major crime category the IVR seemed to perform as well as the CATI version. About one-third of events classified into a TOC were in the wrong category. Once reviewing these incidents, it was found that most of the error was misclassification within the major crime groups of violent and property crimes. For example, a significant number of incidents initially classified as a burglary were actually thefts or motor vehicle thefts. The opposite error also occurred. A large proportion of these errors occurred because respondents did not report a theft when asked. A second type of error occurred because respondents mis-interpreted key questions needed in the TOC algorithm. For example some respondents reported events in the yard as being 'in home', which lead to incorrectly assigning a burglary TOC. Some of this error seemed to be a result of breaking up the questions into a series of 'yes/no' items. As a result, respondents did not hear all of the response alternatives.

Implications. The evidence on overall usability and adapting the instrument provide further evidence that NCVS-1 can be used for the NCVS or a local area survey. We are recommending using a keypad mode of entry to minimize entry errors and levels of frustration on the part of the respondent. The design might also consider giving respondents a choice on which mode to use, however the design of the survey should be driven by the keypad entry mode. It should be possible to incorporate speech for the IVR as the technology in speech recognition advances.

The evidence seemed to indicate a tendency of IVR respondents to report more crimes per person. The prevalence rates were very similar between CATI and IVR, but the number of victimizations reported was higher for IVR. This raises the possibility that the NCVS-1 reports from the IVR

contain ineligible events. One change the design of the IVR that should be considered is to have a more explicit confirmation of the month/year of each incident reported. It would be preferable to collect the month and year for each incident at the time the incident is initially reported on NCVS-1. This will force the respondent to think about the timing of the events relative to the reference period. Additional verification at the screener will increase the chances that respondents will report on an eligible event. For example, verifying the identity of the victim of the incident would prevent respondents from reporting other types of ineligible events.

Use of the IVR for NCVS-2 will involve adapting the instrument more explicitly to a keypad mode. The IVR for this project was designed to accommodate both keypad and speech. All questions were converted to yes/no items to ensure the speech would have minimal problems recognizing utterances. This format was not ideal to adapting critical open-ended NCVS-2 items, such as location and type of items stolen. One adaptation should be to translate the current open-ended NCVS-2 items to multiple choice questions, which can be answered with numbers on the keypad. This would allow respondents to hear all of the response categories before answering. A second adaptation would be to use open-ended, verbatim, questions whenever possible. For example, asking about which items were stolen. Use of this type of question would reduce the overall burden of the instrument.

Question 3: What are the prevalence and victimization rates for the IVR?

This question involved estimating the prevalence and victimization rates with the IVR and comparing these rates to the NCVS, across the experimental conditions and to the CATI.

Comparison to the NCVS. The rates from the IVR are high relative to those produced by the NCVS. The biggest difference is for property crimes, where the IVR rate is almost twice as high as the NCVS. This difference persists even after making a crude adjustment for bounding on the IVR. Given the large number of differences between the two surveys, it is impossible to pinpoint why they differ. However, these results are consistent with several other surveys that have applied the NCVS methodology (Westat, 2013; Biderman, et al., 1985).

Comparison of experimental conditions The experimental treatments that had the highest response rates (ie., mail mode of contact; incentive; keypad) also had nominally higher rates of victimization. However these differences were relatively small and only one was statistically significant (property crime for mode of contact).

Comparison to CATI The IVR did not differ from the CATI with respect to prevalence rates. The rates were almost identical for violent and property crimes. The victimization rates also did not significantly differ when relying solely on the incident forms that were completed. After adjusting for missing incident forms, the IVR victimization rates were consistently higher than the CATI. These differences were only statistically significant when weighting to fully account for all screener reports of incidents. If the weight is capped at six or if one just includes the general population sample, none of the differences are statistically significant.

Implications. There was inconsistent evidence of a mode effect when comparing the IVR to CATI. There was no evidence when examining prevalence rates, while there was a tendency for the IVR to yield more reports of multiple victimization. This may partly reflect the unverified nature of the IVR reports. Whether it was because the incident was capped or the respondent dropped out of the survey, these incidents did not go through NCVS-2 and may disproportionately represent an eligible crime. If the NCVS-2 is to be adopted for the larger NCVS, we recommend further review of the incident summaries taken at the screener in order to assess the quality of the data.

General Recommendations and Next Steps

The two proposed uses of the IVR are: 1) to follow-up respondents after the initial household visit and/or 2) to conduct a local area survey. In both cases, the IVR could be used in two different ways. One would be administer NCVS-1 to screen for victims. The second is to administer both NCVS-1 and NCVS-2 to estimate victimization rates.

Use of IVR as a Follow-up to the Ongoing NCVS. This project provides evidence that an IVR version of NCVS-1 could be used as a way to follow-up respondents after the first NCVS interview. Depending on how many respondents use the IVR, this could reduce the involvement of the interviewer by 20% to 30%. If the NCVS-2 is also administered by IVR, then additional interviewer time would be saved.

To further assess the feasibility of the IVR for the NCVS, additional research should be completed. First, research should further explore the effects of mode on reports of victimization. A relatively low cost investment would be to review the IVR summaries on NCVS-1 from this project to further describe the types of events that were reported on the IVR screener. A second line of research would be to experiment with adapting the NCVS more specifically for this mode of interviewing.

More generally, when considering any self-administered mode within the ongoing NCVS, some investigation of the mode effects needs to be done within the context of the NCVS. The current study was a comparison with a one-time CATI survey. How CATI and IVR compare when the interview is conducted by an experienced NCVS interviewer, either over the telephone or in-person, should be examined. Similarly, if other self-administered modes are being considered, such as the Web or audio computer assisted self interviews (ACASI), then comparisons to these modes should also be completed.

The mode effects should be examined when the survey is used as a follow-up of respondents who have already been interviewed. The present study was conducted as a one-contact design. NCVS respondents generally report fewer incidents at higher times-in-sample. While the reasons for this decrease are unclear, some of the change is due to respondents learning more about the survey and what it covers. Interviewers may also change their behavior at later times-in-sample. The assessment of mode effects in the present study does not take this into account.

In conjunction with this methodological research, more specific cost models should be developed to better understand the cost savings that might be realized with different types of designs. The hypothetical cost savings discussed in this report suggest that significant savings can be achieved. But these estimates were not based on actual costs of the ongoing NCVS nor did it put the IVR within a specific set of design parameters. With more information on each of these, it should be possible to get a better sense of the cost benefits of the IVR for the NCVS.

Use of IVR for Local Surveys. The main advantage of the IVR for a local area is that it is relatively inexpensive to administer. Relative to a web or CATI survey, the IVR does require more specialized technical expertise. If local area agencies were to use the IVR, they would need fairly detailed programming specifications to maximize data quality, as well as to control development costs. Given this, it is not clear whether an IVR would be less expensive for a local agency than other types of surveys that can draw on more familiar survey procedures, such as a CATI survey. If an IVR were to be a viable option for a local area, it would be beneficial for BJS to provide the programming specifications for the basic survey. The local area could modify these specifications to meet their own needs.

This study found the victimization rates for the IVR were somewhat higher than for a CATI application. If BJS is interested in offering IVR to local agencies, the research noted above should be carried out. In particular, the design of the IVR should be adapted to a keypad mode of entry

and more research should be conducted on why IVR respondents seemed to report more victimization when compared to respondents to the CATI.

Limitations of the Study

The limitations of this study are linked to the sample frame used, the relatively small sample sizes for the telephone modes, and the measures of data quality for the NCVS-2 and the length of NCVS-2.

The study used an address based samples (ABS) that were reversed matched to a telephone number. Consequently the results cannot be directly generalized to the entire population of interest to the NCVS. Residents of households with an address that can be matched to a phone number tend to be more cooperative on surveys.

The comparisons across the different modes of contact and interviewing are based on relatively small sample sizes. For example, the CATI completed approximately 280 interviews, while the IVR Telephone interviews completed approximately twice as many. For survey-related outcomes, such as prevalence and victimization rates, statistical tests comparing these to the IVR Mail treatments have relatively low power. This biases the results against finding significant mode differences.

The third limitation of the study is related the review of the summaries taken from the end of the incident form. The review of narrative summaries were all done by one individual who was experienced with the TOC assignments. The coding was limited by the information that was available in the summaries. It was also limited by the use of a single coder, without any independent coding of a subset of the summaries.

Finally, the IVR on this study modified the NCVS design in ways to make it more compatible for this mode. This generally involved shortening the survey and making the questions simpler and easy to provide a response. Wrapped in these adaptations is shortening the NCVS-2 to those questions that were needed to classify the incident. The results reported above are for this shortened version of the NCVS.

Table of Contents

Chapter			Page
Executive Summary			**i**
1.	**Background**		**1-1**
	1.1	Research Questions	1-2
2.	**Methodology**		**2-1**
	2.1	Experimental Design	2-1
	2.2	Sample	2-3
		2.2.1 Sampling of USPS Residential Addresses	2-3
		2.2.2 Sampling of Police-Department-Provided Addresses	2-4
		2.2.3 Creation of Mail-Contact Release Groups	2-6
		2.2.4 Final Sample Sizes	2-6
	2.3	Data Collection Procedures	2-6
		2.3.1 Telephone Initial Contact – CATI Interview	2-6
		2.3.2 Telephone Initial Contact – IVR Telephone	2-7
		2.3.3 Mail Initial Contact – IVR Mail	2-7
		2.3.4 Field Period	2-8
	2.4	Adaptations to IVR Administration	2-8
		2.4.1 Design Principles	2-9
		2.4.2 Question Administration	2-10
		2.4.2.1 Other Changes to NCVS-1	2-10
		2.4.2.2 Other NCVS-2 Changes	2-13
		2.4.3 Link Between NCVS 1 & NCVS 2	2-14
		2.4.4 Other Design Features for IVR Administration	2-15
		2.4.4.1 Error Handling	2-15
		2.4.4.2 Barge-in	2-16
		2.4.4.3 End of Response Detection	2-17
3.	**Response Rates**		**3-1**
	3.1	NCVS-1 Response Rates by Mode and Sample Frame	3-2
		3.1.1 Sample Frame Response Rates	3-4
		3.1.2 NCVS-1 Response Rates and Mode of Response Entry	3-6
		3.1.3 Breakoffs on NCVS-1	3-7
	3.2	NCVS-1 Respondent Characteristics	3-8
	3.3	Non-Response on NCVS-2	3-14
	3.4	Summary and Implications	3-16
4.	**Encouraging Response & Enhancements to Increase Response**		**4-1**
	4.1	Incentive	4-2
	4.2	Mail Insert	4-3
	4.3	Incentive and Demographic Composition	4-4
	4.4	Incentive and the Detailed Incident Form	4-5
	4.5	Summary and Implications	4-5
5.	**Usability**		**5-1**

5.1 Length of the Interview 5-2
5.2 Problems Entering Responses 5-4
5.3 Respondent Perceptions 5-8
5.4 Other Issues 5-10
5.5 Summary of Usability of IVR 5-12
6. Adapting the NCVS-1 6-1
6.1 Response to the Screener 6-2
6.2 Unduplication Procedure 6-3
6.3 Comparison of IVR to CATI 6-5
6.4 Number of Victimizations Reported on NCVS-1 6-6
6.4 Verifying Incidents and Transitioning to NCVS-2 6-8
6.5 Summary of Adapting NCVS-1 for IVR 6-9
7. Adapting NCVS-2 7-1
7.1 Capping the Number of Incidents 7-2
7.2 Review of the Detailed Incident Forms 7-3
7.3 Final Classification and Comparison to CATI 7-9
7.4 Discussion 7-11
8. Comparing Crime Rates 8-1
8.1 Methods for Computing Rates 8-2
8.2 Overall Estimates and Comparisons to the NCVS 8-2
8.3 Comparison of Estimates by Mode of Contact, Incentive Treatment and Mode of entry 8-4
8.4 Comparison to CATI 8-5
8.5 Discussion 8-8
9. Costs of the IVR 9-1
9.1 Cost Assumptions 9-1
9.2 Costs 9-2
9.3 Discussion 9-3
10. Discussion and Recommendations 10-1
10.1 Research Questions 10-1
10.1.1 What are the Response Rates for the IVR? 10-5
10.1.2 Can the NCVS be Adapted for IVR Administration? 10-6
10.1.3 What are the Victimization Rates for the IVR? 10-9
10.2 General Recommendations and Next Steps 10-11
10.2.1 Use of IVR as a Follow-Up to the Ongoing NCVS 10-11
10.2.2 Use of IVR for Local Surveys 10-14
10.3 Limitations of the Study 10-14
R. References R-1

Appendix **Page**

A. Letters Sent to Respondents A-1
B. IVR Questionnaire B-1
C. CATI Questionnaire C-1
D. Priorities When Capping to 3 Incidents D-1

Tables		Page
A.	Experimental Variations Tested on Project	ii
2-1.	Experimental Variations	2-2
2-2.	Processing of USPS addresses	2-3
2-3.	Processing of Police Department addresses	2-4
2-4.	Number of addresses in the CATI-contact samples by experimental group	2-4
2-5.	Number of addresses in the mail-contact samples by experimental group	2-5
3-1.	Detailed Frequency of Final Dispositions at the Household Screening Stage for Each Interview Mode Contacted by Telephone by Sample Frame	3-3
3-2.	Detailed Frequency of Final Dispositions for the NCVS-1 for Each Mode of Contact and Interview by Sample Frame	3-3
3-3.	Response Rates by Mode of Contact and Interview	3-4
3-4.	NCVS-1 Response Rates for Each Sample Group by Mode of Contact, Interview and Metro Area	3-5
3-5.	NCVS-1 Response Rates by Mode of Response for IVR Mail and IVR Telephone Groups	3-7
3-6.	Distribution of Breakoffs within NCVS 1 Interview by Mode*	3-8
3-7.	Household Size for Each Sample Group by Mode of Contact and Interview	3-9
3-8.	Respondent Characteristics for Each Sample Group by Mode of Contact and Interview	3-10
3-9.	Respondent Education for Each Sample Group by Mode of Contact and Interview	3-11
3-10.	Housing Characteristics for Each Sample Group by Mode of Contact and Interview	3-11
3-11.	Household Income for Each Sample Group by Mode of Contact and Interview1	3-12
3-12.	Comparison of ACS and NCVS IVR ABS respondent characteristics	3-14
3-13.	Comparison of total response rates by Mode of Contact and Mode of Interview	3-14
3-14.	Percent of Missing NCVS-2 forms by Mode of entry	3-16
4-1.	Response Rates by Incentive Condition for IVR Mail and IVR Telephone Groups by Sample Frame	4-2
4-2.	Response Rates by Incentive and Insert for IVR Mail Group	4-3
4-3.	Percent of Missing NCVS-2 forms by Incentive Condition	4-5
5-1.	Average time to complete subsections of the IVR by Mode of entry	5-2
5-2.	Average time to complete NCVS sections by Victimization Status and Mode of entry	5-3
5-3.	Proportion of cases experiencing errors (time out / invalid response) in NCVS 1 by completion status	5-5
5-4.	Proportion of cases experiencing errors (time out / invalid response) in NCVS 1 by response mode and completion status	5-6

5-5. Proportion of cases experiencing errors and request for help in NCVS 1 by error type and response mode....5-6
5-6. Proportion of cases experiencing errors (time out / invalid response) in NCVS 2 by response mode....5-7
5-7. Perceived satisfaction with IVR questionnaire....5-8
5-8. Perceived difficulty understanding the IVR system....5-8
5-9. Perceived difficulty of IVR system understanding the respondent (Speech only)....5-9
5-10. Perceived difficulty responding using telephone keypad (keypad only)....5-9
5-11. Perceived difficulty of confusion with IVR questionnaire....5-10
5-12. Perceived amount of time to respond....5-10
6-1. Percent of IVR Respondents Reporting a Crime at the Screener by Screener Item and Sample Type....6-2
6-2. Percent of IVR Respondents Reporting a Crime at the Screener by Screener Item and Mode of Contact....6-3
6-3. Mean number of victimizations reported on the IVR screener by Mode of Contact and Sample Frame....6-7
6-4. Initial number of reports of Victimizations on IVR for NCVS-1 by Number Reported after verification (Percent of those initially reporting a victimization)....6-9
7-1. Percent Distribution of Major Types of Crimes for IVR....7-3
7-2. Percent Distribution of Quality Codes Assigned to Incidents by whether the incident was assigned to a 2-digit TOC code....7-4
7-3. Percent of Original TOCs assigned in each Revised TOC for misclassified incidents....7-8
7-3. Percent Distribution of Types of Crimes Pre- and Post TOC editing....7-9
8-1. Percent of IVR Respondents Reporting at Least One Crime by Type of Crime.....8-3
8-2. Victimization rates for IVR Respondents by Type of Crime.*....8-3
8-3. Percent of Persons Reporting a Victimzation by IVR Experimental Treatments....8-5
8-4. Victimization Rates per 1000 persons by IVR Experimental Treatments....8-5
8-5. Percent reporting at least one victimization persons by mode of interview....8-6
8-6. Victimization Rates per 1000 persons by mode of interview and type of weight....8-6
8-7. Victimization Rates per 1000 persons for ABS sample by mode of interview and type of weight....8-7
9-2. Estimated Cost Per Completed Interview....9-2

Figures **Page**

3-1. Percent of Persons completing all NCVS-2 forms by the number of forms expected....3-15
4-1. Response Rates by Incentive and Insert Experimental Groups, for Keypad and Speech Response Modes (ABS Sample Frame)....4-4

6-1. Mean number of screener items before and after unduplication by Sample Frame....6-5
6-2. Proportion reporting a crime on NCVS-1 by Mode of Interview* 6-6
6-3. Proportion Reporting 1 or More Victimizations by Number Reported and Frame....6-7
6-4. Mean number of Victimizations Reported on NCVS-1 by Mode of interview and Sample Frame+....6-8
7-1. Reasons why Incidents Were Not Classified into an eligible TOC code (n=232)....7-6
7-2. Comparison of Quality Coding and TOC classification by Mode of Interview.... 7-10
8-1. Comparison of IVR for ABS Sample to NCVS 2011 Before and After Adjustment for Unbounded IVR Interviews+....8-4
9-1. Labor and Other Direct Cost Components for estimates of IVR Mail Procedure....9-2

Exhibits **Page**

2-1. NCVS-1 Question 36a with Subparts (a) through (h).... 2-11
2-2. IVR Script for the NCVS-1 Crime Screening Questions 2-12

Background 1

Chapter Highlights and Key Findings

- The NCVS can benefit from the use of computerized self-administered surveys, such as Interactive Voice Response (IVR). It reduces interviewer variance, promotes confidentiality and reduces costs..
- There are two ways IVR can be used as part of the NCVS: 1) as a way to administer the survey after the first visit and 2) as part of local area surveys. The IVR can be used to screen for victimizations or to administer both NCVS-1 and NCVS-2.
- This project addressed three research questions: 1) . What are the response rates with IVR?, 2) Can the NCVS questionnaire be adapted for IVR administration?, 3) What are the prevalence and victimization rates for an IVR administered NCVS?

The National Crime Victimization Survey (NCVS) is the Nation's primary source of information on criminal victimization. Each year data are obtained from a nationally representative sample of the non-institutional population 12 years of age and older. The information collected includes the frequency, characteristics and consequences of criminal victimization in the United States. The survey enables the Bureau of Justice Statistics (BJS) to estimate the likelihood of victimization in the form of rape, sexual assault, robbery, assault, theft, burglary, and motor vehicle theft. The survey provides detailed information on the characteristics of victimization as well as the characteristics of victims. The NCVS provides the largest national forum for victims to describe the impact of crime and characteristics of offenders.

Currently, the NCVS relies exclusively on interviewer-administered modes of data collection. Since the redesign of the early 1990s, these modes have included in-person and telephone interviewing. Until fairly recently, the in-person interviews were completed using paper and pencil instruments. Within the last few years, the survey has converted to field-based, computer assisted personal interviews (CAPI).

The objective of this project is to examine the use of Interactive Voice Response (IVR) as a mode of data collection for the NCVS. There are several possible applications of IVR for the NCVS. As part of the NCVS core methodology, an IVR mode could be incorporated as part of a multi-mode design within the rotating panel design. For instance, after the initial in-person interview, respondents could

be asked to call into an 800 number to complete the survey in subsequent contacts. By establishing rapport at the first interview, the NCVS could efficiently collect data from some portion of the sample without incurring the expense of an interviewer-driven methodology. A second possible use of IVR is for a supplemental survey to generate local area estimates. Previously, agencies have relied on mail or telephone surveys to conduct local area victimization surveys. IVR could provide a way to increase the efficiency, and possibly the quality, of these surveys.

1.1 Research Questions

This project is designed to examine three basic questions:

1. What are the response rates with IVR?
    a. What are the response rates using inbound and outbound contact methodologies (mail contact vs. telephone contact)?
    b. What is the composition (e.g., socio-demographics) of those that respond to the IVR and how does this differ by mode of initial contact?
    c. Is it possible to effectively encourage sampled households to complete the IVR interview
2. Can the NCVS questionnaire be adapted for IVR administration?
    a. Are respondents able to complete the two major components of the NCVS (NCVS-1, NCVS-2)?
    b. Can the IVR handle tasks normally done by the interviewer, such unduplicating incidents and transitioning to the NCVS-2
    c. What is the quality of the incident data from the IVR?
    d. Are there differences in respondent acceptance between speech IVR and touchtone data entry (Keypad)?
3. What are the Victimization Rates for the IVR?
    a. Does IVR lead to different victimization rates from a telephone interview?
    b. Is there a difference in victimization rates for Speech and Keypad modes of entry?
    c. Is there a difference in victimization rates by method of contact or incentive?

The above research questions speak to two important issues related to the IVR. One is whether it is possible to successfully recruit a general population sample to take the IVR and the conditions under which this recruitment is most successful (question 1). The project evaluated what the overall response rate is for the different methods of recruitment and response. It also provides data on whether particular enhancements (e.g., inserts; promised incentives; use of CATI to recruit) significantly boost the response rate. This information is particularly useful for applications that might use the IVR as a stand-alone methodology, such as boosting sample for local area estimates. Previous locally generated victimization surveys have used simplified methodologies that may not be totally comparable to the NCVS. An IVR application can potentially maintain some of the complexities associated with the NCVS instrument (e.g., screening, detailed incident form), without an interviewer being involved.

The second type of research question relates to the usability, adaptability and measurement properties of the IVR for the NCVS interview (e.g., questions 2 and 3). If an IVR were to be used as either a stand-alone vehicle or as a way to follow-up respondents who are already in the NCVS panel, it is important to assess whether IVR can be adapted to the relatively complicated interviewing protocols used on the NCVS. The NCVS procedures are dependent on the interviewer to make a number of judgments, such as when to accept a report during the screening, how to define incidents for administration of the detailed incident form and determining the nature of the event (e.g., administration of series crimes; location of the event; attempts vs. completed; defining sexual assault). One simplified application of IVR would have respondents complete the crime screener to find out if the respondent has been victimized during the reference period. For those that report a victimization, the interviewer could follow-up with those that report any victimization or a subsample of those individuals by administering the detailed incident form. This could be applied at panel waves after the first interview or as part of a screening exercise associated with a local survey. A more complicated application would ask respondents to fill out both components of NCVS, the crime screener (NCVS 1) and the detailed incident form(s) (NCVS 2). The purpose of the present project was to assess the utility of the IVR for each of these types of applications.

Methodology 2

Chapter Highlights and Key Findings

- The project carried out an experimental design that compared: 1) interviewer vs. IVR mode of administration, 2) inbound (mail) and outbound (telephone) mode of contact for the IVR, 3) method of entry for the IVR (speech vs. Keypad) and 4) methods to enhance IVR response rates.
- The sample was carried out in two cities using two sample frames: 1) USPS postal delivery addresses (ABS) and 2) addresses of households where someone had reported a crime to the police in each central city. Most of the sample consisted of addresses where a match to a telephone number could be obtained.
- To administer the NCVS on the IVR, a number of adaptations were made, including: 1) modifying the procedures to administer NCVS-1, 2) reducing the length of NCVS-2 and 3) converting open-ended items to a series of 'yes/no' questions.

This chapter describes the methodology used for the field experiment. The first section describes the experimental design. The second section describes the sample design. The third describes the procedures used to make contact with respondents and the fourth describes the design of the IVR and CATI versions of the questionnaire that were used in the project.

2.1 Experimental Design

The project implemented a design which experimentally varied a number of different factors, including the method of contact (telephone interviewer vs. mail), mode of response (CATI vs. IVR; IVR-speech vs. IVR-touchtone), and methods to enhance the response rate (promised incentive vs. no incentive; insert vs. no-insert). Table 2-1 provides these variations.

Two methods of contact were used for the IVR: 1) using a telephone interviewer (IVR Telephone) and 2) mail contact through the use of an invitation letter (IVR Mail). Participants contacted by telephone were transferred to the IVR system after a respondent had been selected. For mail contact, the invitation letter instructs the person with the next birthday to contact the IVR. A

telephone interview mode is also included for comparison where respondents complete the NCVS components in CATI.

The IVR system varied the mode of response, with touchtone data entry (Keypad) and speech response modes. Households were randomly assigned to either Keypad or speech response. Keypad uses the telephone keypad where instruction is provided on the appropriate keys for response. Speech requires the respondent to speak their response. In the IVR application, all speech responses were restricted to 'yes' or 'no' whenever possible. Restricting the vocabulary in this way made it easier to program the speech application. For all questions in the interview that required a numeric response, respondents assigned to either mode of entry defaulted to Keypad entry. Reasons for this are explained later.

Two methods were included to enhance response rates; the offer of a promised incentive, and inclusion of an insert with the mail invitation letter. Half of the respondents in the IVR Mail and IVR Telephone modes of entry were randomly assigned to receive a $10 monetary incentive after the completion of the NCVS components of the interview. Households contacted by telephone were told of the promised incentive after a respondent was selected, but just prior to the transfer to the IVR interview. They were then informed of the promised incentive again during the IVR introductory text. Households contacted by mail were informed about the incentive in the body of the invitation letter, and again during the IVR introductory text.

Households in the IVR Mail group were randomly assigned to either an insert or no insert condition. The insert included a motivating statement about the survey, and included the telephone number and ID needed to access their IVR survey. For households assigned to the incentive group, the mention of the promised incentive appeared on the insert in lieu of the statement.

Table 2-1. Experimental Variations

		Response Entry Method			
		Speech		Keypad	
IVR Mail (Mail Contact)		$0	$10	$0	$10
Insert	Yes				
	No				
IVR Telephone (Phone Contact)					
CATI (Phone Interview)					

2.2 Sample

The survey was carried out in the St. Louis and Houston metropolitan statistical areas. There were two sample frames: 1) all residential addresses based on United States Postal Service (USPS) administrative data and 2) addresses where someone had reported a crime to the St. Louis and Houston police departments within the last 12 months. The two sources of addresses were each used to produce 13 independent samples of addresses in each community. In the next sections we discuss separately the sampling of USPS residential addresses and police department addresses.

2.2.1 Sampling of USPS Residential Addresses

The first step in sampling the USPS residential addresses was to subset them. This was done by using Arc Info geographical-information-system software to delineate each community's urban areas of the two cities (St. Louis and Houston). The sampling frame of USPS residential addresses was then subsetted to those in 2000 Census tracts for which the percentage of the tract in the delineated urban area was equal to greater than 50 percent. The subsetted USPS residential addresses were then sorted by Zip+4 codes and carrier routes, and 22,500 addresses were randomly selected for each community using equal-probability systematic sampling with a random start.

In order to maintain comparability between the mail contact and telephone contact methods, only addresses for which a telephone number could be found were sampled. The sampled USPS residential addresses were reverse matched to telephone numbers. The match rate was 45.7 percent in Houston and was 58.4 percent in St. Louis. For each community, the matched addresses were used to create five random samples of addresses in which outbound CATI was used to initiate respondent contact and another eight random samples in which mail was used to initiate contact. Table 2-2 provides the result of this process.

Table 2-2. Processing of USPS addresses

	Houston	St. Louis	Both
First-phase sample	22,500	22,500	45,000
Match rate	45.7%	58.4%	
Matched addresses	10,275	13,130	23,405
Assigned for CATI contact	1,669	1,746	3,415
Assigned for mail contact	6.313	6,600	12,913
Not Assigned	2,293	4,784	7,077

The five CATI-contact samples were used to allocate to the five different experimental groups. The eight mail-contact samples were used to compare different protocols for using mail-invited IVR.

2.2.2 Sampling of Police-Department-Provided Addresses

Similar to the sampling of the USPS residential addresses, the addresses for the police department addresses were reverse matched to find a telephone number. The match rate was 39.4 percent for Houston and 58.9 percent for St. Louis. The matched addresses were used to create five CATI-contact samples, one for each experimental group. The matched addresses were also used to create eight mail-contact samples for each community (Table 2-3). Column 5 of Table 2-4 contain the number of matched addresses in the CATI-contact samples of police department addresses. For the mail-contacts, some of the unmatched addresses were retained in order to get an estimate of the response rate for these types of households. Columns 5 and 6 of Table 2-5 contain the number of matched addresses and the number of unmatched addresses, respectively, in the mail-contact samples of police department addresses.

Table 2-3. Processing of Police Department addresses

	Houston	St. Louis	Both
Provided addresses	5,289	7,745	13,034
Match rate	39.4%	58.9%	
Matched addresses	2,085	4,563	6,648
Fielded for CATI contact	341	744	1,085
Fielded for mail contact	1,237	2,547	3,784
Not fielded	507	1,272	1,779
Unmatched addresses	3,204	3,182	6,386
Fielded for mail contact	1,505	1,752	3,257
Not fielded	1,699	1,430	3,129

Table 2-4. Number of addresses in the CATI-contact samples by experimental group

	Treatment		Address source and reverse-match status		
Community	Interview Completion	Incentive?	USPS, matched	Police, matched	Total
Houston	Interviewer		557	113	670
	Voice IVR	Yes	278	57	335
		No	278	57	335
	Keypad IVR	Yes	278	57	335
		No	278	57	335

Table 2-4. Number of addresses in the CATI-contact samples by experimental group (Continued)

Community	Treatment		Address source and reverse-match status		
	Interview Completion	Incentive?	USPS, matched	Police, matched	Total
St. Louis	Interviewer		582	248	830
	Voice IVR	Yes	291	124	415
		No	291	124	415
	Keypad IVR	Yes	291	124	415
		No	291	124	415
Both	Interviewer		1,139	361	1,500
	Voice IVR	Yes	569	181	750
		No	569	181	750
	Keypad IVR	Yes	569	181	750
		No	569	181	750
Both-TOTAL			3,415	1,085	4,500

Table 2-5. Number of addresses in the mail-contact samples by experimental group

Community		Treatment		Address source and reverse-match status			
	IVR type	Insert?	Incentive?	USPS, matched	Police, matched	Police, not matched	Total
Houston	Voice IVR	Yes	Yes	789	157	189	1,135
			No	790	152	193	1,135
		No	Yes	789	156	187	1,132
			No	789	166	178	1,133
	Keypad IVR	Yes	Yes	789	158	184	1,131
			No	789	151	193	1,133
		No	Yes	789	148	192	1,129
			No	789	149	189	1,127
St. Louis	Voice IVR	Yes	Yes	825	335	203	1,363
			No	825	306	231	1,362
		No	Yes	825	322	215	1,362
			No	825	340	197	1,362
	Keypad IVR	Yes	Yes	825	307	231	1,363
			No	825	306	231	1,362
		No	Yes	825	321	216	1,362
			No	825	310	228	1,363
Both	Voice IVR	Yes	Yes	1,614	492	392	2,498
			No	1,615	458	424	2,497
		No	Yes	1,614	478	402	2,494
			No	1,614	506	375	2,495
	Keypad IVR	Yes	Yes	1,614	465	415	2,494
			No	1,614	457	424	2,495
		No	Yes	1,614	469	408	2,491
			No	1,614	459	417	2,490
Both-TOTAL				12,913	3,784	3,257	19,954

2.2.3 Creation of Mail-Contact Release Groups

The mail-contact sampled addresses were randomly assigned to 10 releases groups, which provided the capability to field some or all of the mail-contact sampled addresses. The first release group contained approximately 28 percent of the mail-contact addresses. The other nine release groups each contained approximately nine percent of the addresses.

2.2.4 Final Sample Sizes

All 4,500 addresses for the CATI contact sample were released for the study. Only the first four release groups for the Mail IVR were mailed. This resulted in a final sample size of 9987 for the Mail IVR group. Consequently, the final sample released for the various treatments was 14,487.

2.3 Data Collection Procedures

The data collection procedures for each group differed in both how they were contacted and whether they were interviewed by a telephone interviewer, or whether the questions were administered by the IVR system. Contact materials are provided in appendix A.

2.3.1 Telephone Initial Contact – CATI Interview

The CATI group's initial contact was by a telephone interviewer. The telephone interviewer administered a short household screening questionnaire to identify and sample one household member. The household member was sampled by asking for the person with the next birthday. The selected household member was then administered the crime screener (NCVS-1) and detailed incident form (NCVS-2) by the telephone interviewer using the CATI system. All households in this group were mailed an initial pre-notification letter informing the household they would be contacted by a telephone interviewer. A $2 cash incentive was included in the prenotification letter mailed to this group.

2.3.2 Telephone Initial Contact – IVR Telephone

The IVR Telephone group's initial contact was identical to the CATI interview. The telephone interviewer administered a short screening questionnaire to identify and sample one household member. The respondent was selected by asking for the person with the next birthday. Once the interviewer had the selected household member on the telephone, the interviewer transferred the selected respondent to the IVR system to administer the NCVS-1 and NCVS-2 (if needed). Prior to the transfer to the IVR system respondents were informed they would be transferred to an automated system and what mode of response they would be expected to use (Keypad or Speech).

If the sampled respondent did not complete the IVR interview during the phone contact (e.g., hung up before finishing the IVR) they were placed back into the queue of respondents to be re-contacted by the CATI telephone interviewer.

All households in this group were mailed an initial prenotification letter informing the household they would be contacted by a telephone interviewer. This letter was identical to the letter mailed to the CATI group. A $2 cash incentive was included in the prenotification letter mailed to this group.

2.3.3 Mail Initial Contact – IVR Mail

The mail contacts for this group generally followed procedures established for mail surveys. All addresses were mailed an initial invitation letter, with the toll-free number to contact the IVR system and their ID prominently displayed within the letter. A $2 cash incentive was included in the initial mailing. The letter asked for the adult with the next birthday to call the toll-free number and complete a computerized interview. A thank-you/reminder postcard was sent to all households about one week after the initial mailing. The postcard included the toll-free number for contacting the IVR system and the eight digit ID to access their survey. Two weeks later all non-responding addresses including all addresses that had accessed, but not completed the IVR survey, were mailed a follow-up letter. This letter again invited them to complete the IVR survey and included the same toll-free number and eight digit ID. Two-weeks after the first follow-up all nonresponding addresses at this point were mailed a final follow-up letter. The letter included the same toll-free number and ID number as previous letters.

The first two mailings used a standard number 10 envelope. The final nonresponse mailing used a larger 9" x 12" envelope to increase the salience of the mailing. As noted when describing the

experimental design, the IVR Mail sample experimented with the use of an insert which had a short message intended to motivate the respondent to complete the survey.

2.3.4 Field Period

Data collection for the two groups whose initial contact was by telephone interviewer began on 2/27/2012 and concluded 4/19/2012. The prenotification mailings for this group were mailed on 2/23/2012.

For the IVR Mail group, the first wave was mailed on 3/7/2012 and the last follow-up of the fourth wave was mailed on 5/18/2012. The IVR system was active and accessible starting on 2/27/2012 through 6/4/2012. After 6/4/2012 anyone attempting to contact the IVR system would hear a message that the study had concluded, and thanking them for their interest.

Data collection for the IVR Mail group was delayed due to an issue relating to missing apartment numbers from one of the police departments. This arose during the process of telephone matching where a number of cases within the sample had duplicate telephone numbers. The start of the mailings for this portion of the sample was delayed to allow time to resolve this issue.

The sample contacted by telephone was released in a single group. The IVR Mail sample was released in waves. This was done to moderate the load on the IVR system. The IVR system had a limited capacity with respect to the number of calls it could receive at the same time. Releasing the sample in waves controlled this to the extent necessary. Four waves of IVR Mail mailings were used. The first wave included 4,000 addresses, while the second, third, and fourth waves included 2,000 addresses each. The start of each wave was separated by one week with the exception of the second wave, which followed two weeks after the start of the first wave.

2.4 Adaptations to IVR Administration

The NCVS is administered by an interviewer using computer assisted personal interviewing (CAPI). In some cases these interviews are conducted by telephone utilizing the same CAPI instrument. The NCVS Mode Study differed from this process by using a self-administered procedure using IVR. In IVR administration a computer rather than a live interviewer asks each question by playing a series

of recordings over the telephone. These recordings can vary and can incorporate responses to earlier questions to tailor the current question. Respondents provide their answer to each question by one of two pre-assigned modes of response. One method of response is Touch Tone Data Entry (Keypad). Respondents press a key on the telephone keypad that corresponds to their answer. The other method is Speech, in this method respondents say their answer which is recognized by the IVR system.

For an interviewer-administered survey a process known as grounding occurs. In this process the interviewer and respondent engage in back and forth dialogue in an effort to establish a common understanding (Clark & Brennan, 1991). For example, interviewers use this procedure when establishing what a particular question is asking, or clarifying a response. The interviewer can also detect inconsistencies over the course of the interview. Specifically for NCVS, respondents may report a single victimization multiple times. Other examples might include unduplicating incidents that were already reported or clarifying the relevance of particular incidents for the survey. The IVR needed to be adapted to account for this process as much as possible. Appendices B and C provide the IVR and CATI questionnaires.

2.4.1 Design Principles

The adaptations of the NCVS to the IVR focused on several design principles.

- Keep the task simple: IVR respondents must be able to understand the question and eligible responses the first time they hear it. Respondents may not wait to hear a question repeated, or may select any valid response.
- Do not over-burden the respondent: while related to the first principle, this relates to question length and number of response categories to include. A question that is lengthy may lose the attention of the respondent or may overtax their memory. Including too many response categories will similarly lead to overtaxing memory and may lead to response order effects, best-guess responses, or termination of the response task.
- Keep the respondent engaged. Unlike in an interviewer administered interview, the question and response process is unlike what is generally a normal social exchange with another individual. The respondent must 'wait their turn' and may have limited opportunity to interrupt the system. Interviewers may vary the pace of the interview, while an IVR application will proceed at a measured constant pace across all interviews. Minimizing question length and the amount of introductory or instructional text will help keep the respondent engaged in the task. Increasing engagement with the IVR

application will prevent respondents from "drifting off" during the interview, and contribute to the perception of progressing through the interview.

When making changes to the questionnaire, the structure and wording of the NCVS were preserved to the extent possible. All changes were tested through three rounds of usability testing culminating in a dry run pilot test in order to test support systems and procedures.

All adaptations described below are based on the NCVS 1 and NCVS 2 forms used by the U.S. Census Bureau. The versions used here are: NCVS 1, implementation date 09-16-2004; NCVS 2, implementation date 09-16-2004.

2.4.2 Question Administration

One adaptation of the NCVS for the IVR was decomposing questions to simplify the responses required of the respondent. It was also desirable to simplify the response categories to keep the speech mode of entry viable. The most drastic adaptation was to change most questions to a 'yes' or 'no' format. This had more significant implications for NCVS-2, where there are a number of open-ended questions with the interviewer coding the response into a particular category. For example, item 26 on NCVS-2 asks how the respondent was attacked. This open-ended item was broken into a series of 'yes/no' questions using the response categories:

Q26. Did the offender attack you in any of the following ways?
QB. Were you raped? *Expected Response: YES or NO*
QN. Did the offender attack you in any other way? *Expected Response: YES or NO*

A few items have responses that are either clearly implied by the question, or are generally understood. Examples are asking gender, or marital status. For these items, respondents were asked directly for the appropriate response category.

2.4.2.1 Other Changes to NCVS-1

For NCVS-1, almost all of the items were already in a 'yes/no' format. The process of reading the set of cues for each of the victimization screener items was modified. The logic was structured

around the different subquestions associated with each of the victimization screener items. The change can be illustrated with the first crime screener item for theft (Item 36a) shown in Exhibit 2-1.

Exhibit 2-1. NCVS-1 Question 36a with Subparts (a) through (h)

36a. I'm going to read some examples that will give you an idea of the kinds of crimes this study covers.

As I go through them, tell me if any of these happened to you in the last 6 months, that is since _____ _____, 20 ____.

Was something belonging to YOU stolen, such as –

(a) Things that you carry, like luggage, a wallet, purse, briefcase, book –
(b) Clothing, jewelry, or cellphone –
(c) Bicycle or sports equipment –
(d) Things in your home – like a TV, stereo, or tools
(e) Things outside your home such as a garden hose or lawn furniture –
(f) Things belonging to children in the household –
(g) Things from a vehicle, such as a package, groceries, camera, or CDs –

OR

(h) Did anyone ATTEMPT to steal anything belonging to you?

This has eight subparts (36aa through 36ah). Under the interviewer administered version, respondents are read each subpart in one continuous string. If respondent interrupts with a 'yes' to a particular subpart, the interviewer continues to read through the rest of the subparts. At the end of the question, the respondent is then asked to describe each incident and the number of times it occurred.

To maintain more logical flow for the IVR, some of the subparts were turned into 'yes/no' items. In some cases, the screener subparts were combined to reduce the number of 'yes/no' responses the respondent had to provide. The final scripts used for the screener items is shown in Exhibit 2-2. The exhibit provides the NCVS-1 screener number. The subpart questions that have been combined for the IVR are shown in parentheses next to the subpart number. For example, Item 36a combined all of the subparts (a) through (h) into a single question. One subpart was not included at all – this was for item 41a, which excluded the overall 'attempt' subpart. This was done to reduce the redundancy with the immediately prior subpart, which asks about attempted rape.

Exhibit 2-2. IVR Script for the NCVS-1 Crime Screening Questions

NCVS-1: 36a

1. (a-g) "In the last 12 months, that is since {CURRENT MONTH/YEAR – 12 MONTHS}, was something belonging to YOU stolen, such as:
   - things you carry, like luggage, a wallet, purse, briefcase, or books;
   - clothing, jewelry, or cellphones;
   - a bicycle, or sports equipment;
   - things in your home, like a TV, stereo, or tools;
   - things outside your home, such as , a garden hose, or lawn furniture;
   - things belonging to children in the household, or
   - things from a vehicle, such as, a package, groceries, camera, or CD's."

OR

2.(h) "In the last 12 months, did anyone attempt to steal anything belonging to you?" SQTHEFTA

NCVS-1: 37a

1.(a) "Other than any incidents already mentioned, has anyone broken in or attempted to break into your home?" (SQVER1)
2.(b) "Has anyone illegally gotten in or tried to get into a garage, shed, or storage room?"
3.(c) "Has anyone illegally gotten in or tried to get into a hotel or motel room or vacation home where you were staying?"

NCVS-1: 39a

1.(a, b) "During the last 12 months, other than any incidents already mentioned, {was the vehicle / were the vehicles} stolen or used without permission?"
2.(c) "Did anyone steal any parts such as a tire, car stereo, hubcap, or battery, or steal any gas from a vehicle?"
or
3.(d) "Did anyone attempt to steal any vehicle or parts attached to a vehicle?"

NCVS-1: 40a

1.(a, b) "During the last 12 months, other than any incidents already mentioned, were you attacked or threatened or did you have something stolen from you at home or near a friend's, relative's, or neighbor's home?"
2.(c) "Were you attacked or threatened or did you have something stolen from you at work or school?
3.(d, g) "Were you attacked or threatened or did you have something stolen from you in places such as a mall, or restaurant, or other places you go for entertainment or recreation such as a party, theater, or gym?"
4.(e, f) "Were you attacked or threatened or did you have something stolen from you on the street, or while riding in any vehicle?"
Or
5. (h) "Did anyone attempt to attack or attempt to steal anything belonging to you from any of these or other places?"

NCVS-1: 41a

1.(a – c) "Other than any incidents already mentioned, has anyone attacked or threatened you with a weapon, such as a gun or a knife, scissors or anything that could be thrown, such as a rock or a bottle?"
2.(d, f) "Has anyone made any face-to-face threats or attacked or threatened you by grabbing, punching, or choking you?"
3.(e) "Has anyone attacked or threatened you with rape, attempted rape or any other sexual attack?"

NCVS-1: 42a

1.(a-d) "People often don't think of incidents committed by someone they know. Other than any incidents already mentioned, did you have something stolen from you OR were you attacked or threatened by someone at work or school, a neighbor or friend, a relative or family member, or any other person you've met or known?"

Exhibit 2-2. IVR Script for the NCVS-1 Crime Screening Questions (Continued)

NCVS-1: 43a

1.(a – c) "Incidents involving forced or unwanted sexual acts are often difficult to talk about. Other than any incidents already mentioned, have you been forced or coerced to engage in unwanted sexual activity by someone you didn't know before, a casual acquaintance, or someone you know well?"

NCVS-1: 44a

1. "During the last 12 months, other than any incidents already mentioned, did you call the police to report something that happened to you which you thought was a crime?"
(if yes) "Were you attacked or threatened, or was something stolen or an attempt was made to steal something that belonged to you or another household member?"

NCVS-1: 45a

1. "During the last 12 months, other than any incidents already mentioned, did anything which you thought was a crime happen to you, but you did not report to the police?"
(if yes) "Were you attacked or threatened, or was something stolen or an attempt was made to steal something that belonged to you or another household member?"

A second modification to NCVS-1 was to immediately follow-up a 'yes' response to a subpart by asking the number of times the incident occurred and for a verbal description of each incident. Once this was collected, the program went to the next screener item, skipping any remaining subparts for that question. By skipping the other subparts, we were reducing burden on the respondent. However, this does create a difference with the interviewer version, where all subparts were read when a respondent initially responds affirmatively to a particular cue. Presumably this change might depress the victimization rates on the IVR version, assuming that exposure to all of the cues would increase the number of victimization reports. This would only occur for respondents who already have responded affirmatively to subpart to the question. If the respondent reports no to the initial subparts, all cues were administered.

2.4.2.2 Other NCVS-2 Changes

The NCVS-2 posed a different set of challenges when compared to the screener. The need to convert all open ended items to 'yes/no' questions increased the length of the NCVS-2. To reduce burden, the NCVS-2 was reduced to those items that were needed to classify incidents into a type of crime (TOC) code. The only exception to this was to also include the questions on notification of the police. This change was made for both the IVR and CATI versions of the questionnaire.

When converting some of the IVR items to yes/no format, it was possible to create a hierarchical set of items, then limiting the level of detail necessary or the number of responses that would be

asked as a question. One example of this is asking where the victimization occurred. Instead of asking about each listed category, the respondent was first asked the higher level categories and then branched to the more detailed locations within that category.[1]. This strategy was not used where items used by the TOC could potentially be skipped with this change.

Another example is the question asking what was taken from the respondent (Item 96). If the respondent reported a theft they are asked about each higher level category (e.g. cash, vehicles or parts, household furnishings, etc.). If a respondent answered yes to the higher level property category they are asked the detailed response categories as yes or no questions.

2.4.3 Link Between NCVS 1 & NCVS 2

For each affirmative response to a crime item on NCVS 1 respondents were asked the frequency of the victimization. Respondents were then asked to give a verbal description of up to the three most recent victimizations that were recorded by the IVR system. This was repeated for each screener item with an affirmative response.

During usability testing it was observed that respondents were frequently over-reporting victimizations. In many cases the respondents' felt the different victimization questions in NCVS 1 were attempting to collect additional detail about their victimization rather than asking if another type of victimization occurred. To address this, the IVR questionnaire included a verification question for each victimization reported after the first report. The verification question asked if the new victimization they were reporting was the same as one reported earlier. If the response was yes, this victimization was ignored by the system and no frequency or verbal reports were collected.

An NCVS 2 report was created for each victimization reported and confirmed in NCVS 1. Respondents were asked to fill out an incident form for up to three verified incidents. If more than three incidents were reported, three incidents were selected using a predetermined hierarchy (see discussion below).

At the beginning of the NCVS-2, an additional confirmation step was included that verified the frequency and time period of the reported victimization by asking:

[1] The most recent version of NCVS-2 uses an identical strategy (see item 10a on NCVS-2; OMB No. 1121-0111).

"You said that during the last 12 months {TYPE OF INCIDENT} occurred {NUMBER OF TIMES}. Is that correct?"

If the respondent did not confirm this, the system then asked for the new frequency of the victimization. If the respondent reported zero, indicating the victimization did not occur or did not occur within the last 12 months, the victimization was ignored by the system. The system then continued with the next victimization reported, if any. The answers to these confirmation questions would not affect which incidents were asked on the detailed incident form. For example, if a respondent who had three of four incidents selected for NCVS-2 reported a '0' for one of the NCVS-2 incidents, the system would not try to ask for the fourth incident that was initially reported, but not selected. Similarly, if the respondent reported more incidents to the confirmation, this was not used to re-set which incidents would be asked on NCVS-2.

To limit the interview burden experienced by the respondent the number of combined incidents the system would ask was limited to three. More serious and less common incidents were prioritized in a hierarchy. This ensured less common victimizations that were generally more serious would not be omitted from NCVS 2 by more common and less serious victimizations. The victimization hierarchy is shown in appendix D. The prioritization implemented gave preference to different victimization types over multiple incidents. For example, if three thefts and two assaults were reported, the system would include the most recent incident of each victimization type (theft and assault) and one other incidence based on the prioritization.

2.4.4 Other Design Features for IVR Administration

Several design features were incorporated into the IVR questionnaire to handle situations where the respondent does not provide a valid response, or the IVR system is unable to determine the response provided.

2.4.4.1 Error Handling

Whenever the IVR system encounters an instance where the response provided is unrecognizable, invalid, or absent, feedback is provided to the respondent tailored to each type of error. The error types and feedback offered are listed below:

- Invalid keypad response: This may be the result of an inadvertent press of the wrong key, or a 'fat finger'. When this occurs the respondent is told the response they provided is not valid and what their response was in case they are unaware of the error.

- No response: For most questions, respondents were provided six seconds to answer a question. A few questions allowed 10 seconds. When a response was not detected, the respondent is prompted by informing them they have not responded, or in the case of speech respondents, that their utterance was not loud enough. In case a response was not provided because the respondent did not hear the question clearly, the question is repeated.

- Unrecognizable spoken response: In speech mode when the response is indecipherable by the speech recognition system, respondents were informed of this by the IVR system. Reasons for this occurring may be extraneous noise, response disfluencies, additional utterances, or invalid responses. When this occurs the IVR system offers feedback on the expected response, when the response expected for the question is a yes or no response.

2.4.4.2 Barge-in

For interviewer administered surveys the interviewer has the ability to acknowledge interruptions or note when the respondent has completed their response. Examples of interruptions are when the interviewer is reading a long list of examples (e.g. types of things that can be stolen) and the respondent interrupts before the interviewer has completed the list. Interruptions also frequently occur while the interview is reading response categories. For IVR administration the respondent generally must wait until the IVR system has asked the question before it will be ready to accept a response. This can increase the perception that the interview is mechanical and less responsive to the respondent.

The IVR interview allowed for respondents to interrupt, or 'barge-in' with their response once the question began. This allowed respondents to influence the pace of the interview by providing their answer once they felt they heard enough of the question to respond. Most questions were very short, such that, interrupting the system was unnecessary. The most notable exception to this were several of the crime screening items, which contain a number of different cues. For the keypad response mode, instructions on what keys to press on the telephone keypad were provided after each question. Since a large majority of the questions accepted yes or no, this could make the interview feel redundant if respondents were forced to wait until the end of the question. Barge-in functionality allowed the respondent to input their response once they were familiar enough with the pattern.

When response instructions were played (e.g. "for yes press 1, no press 1"), usability testing indicated that it was less confusing to respondents to allow them to barge in, rather than to remove the instructions once they had been heard several times. For the keypad response mode, respondents pick up on the response pattern quickly. However, when these instructions are removed it was not clear to respondents that it was their turn to respond. While the instructions seem repetitive, they served as an indication to the respondent that it is their turn to respond.

2.4.4.3 End of Response Detection

An important feature implemented for open-ended responses is the detection of when the respondent has stopped speaking. The most common example of this is descriptions of victimizations. In interviewer administered surveys the interviewer may note the respondent has finished their response by use of terminating sentences, changes in voice inflection, or other cues. For open-ended responses the IVR system allowed 60 seconds of recording time. Respondents were instructed to press the star (*) key once they were finished to notify the system they were done. However, it was common for respondents to forget (or not know) to press * when finished. In these situations, once the IVR system detected 10 seconds of 'dead air' the system stated that it noted the respondent stopped speaking and reminded them to use the star key next time to indicate when they have finished. Without this functionality respondents may feel the system has failed or broken down in some way if there is still a substantial amount of time remaining before the next IVR action.

Response Rates 3

Chapter Highlights and Key Findings

- The IVR Mail and the CATI interview had equivalent response rates (around 23 percent). The IVR Telephone had the lowest response rate, partially from hang-ups that occurred when transferring to the IVR. These hang-ups were disproportionately concentrated in the speech data entry mode.
- Breakoffs during NCVS-1 were equivalent between the IVR and CATI modes. For both modes, most breakoffs occurred before the victimization questions.
- For the ABS sample, the socio-demographic composition of respondents to each of the modes and methods of data entry were similar. The IVR-Mail and CATI were the closest. The IVR-Telephone had older respondents, respondents who had a 'separated' marital status and did not own their home. All of the methods significantly under-enumerated difficult to several high risk groups (e.g., young, low education, Hispanics).
- Both the IVR-Mail and IVR-Telephone had significant reductions in the overall response rate for NCVS-2. Overall, approximately 30 percent of the IVR respondents did not complete all of their expected NCVS-2 forms. The CATI respondents filled in all eligible NCVS-2s.
- Take-away point #1: The IVR Mail can achieve equivalent response rates and sample composition for NCVS-1 when compared to a CATI survey. At least from a response rate perspective, IVR Mail seems promising for an inexpensive way to screen households for victimization, either for local areas or as part of later waves of the NCVS.
- Take-away point #2. The use of the IVR to collect data on the NCVS-2 is also feasible. However, there will be a significant number of individuals who do not fill out all incident forms. Further experimentation is needed to see if it is possible to reduce the number of who do not fill out all forms by shortening or adapting the NCVS-2 to an IVR mode. As a follow-up mechanism for the ongoing NCVS, this issue could be dealt with through follow-up with respondents after they have completed the survey.

In this chapter we describe the response rates for the IVR interview. The rates for the NCVS-1 and NCVS-2 are discussed separately. Breaking the rates out this way provides an idea of how an IVR interview might work if the goal is to administer the NCVS-1 or both the NCVS-1 and NCVS-2.

3.1 NCVS-1 Response Rates by Mode and Sample Frame

Response rates were decomposed by the different stages of the interview. For the IVR Mail group, there was one stage— NCVS-1. For households contacted by telephone (CATI and IVR Telephone) two stages are shown. First a household screener had to be completed in order to: 1) confirm the telephone was for a residential household and 2) select an adult to complete the interview. Second, the NCVS-1 was administered to the selected respondents. Any adult household member could complete the household screener, but only the selected household member could complete the NCVS. For both the mail and telephone contacts, the adult with the next birthday was selected as the NCVS respondent.

For the telephone contacts, the address was confirmed at the beginning of the call. All cases were considered eligible regardless of address match status. Overall, 92.4 percent of completed household screeners matched the address, 6.4 percent did not match and 1.1 percent refused to confirm.

For the IVR Mail group, postal mail returns that identified non-residential, undeliverable, or vacant are counted as ineligible and excluded from response rate calculations. For the households contacted by telephone, those identified as non-residential and non-working numbers are counted as ineligible.

Response rates were calculated using AAPOR RR1 response rate formula. This is the most conservative method, since it counts all 'unknown' cases as a non-response. To count as a completed survey, the respondent had to finish all of the NCVS-1. For purposes of comparison, the IVR cases exclude those addresses where a promised incentive was offered.

Table 3-1 provides a frequency of the detailed results for the household screening stage for the telephone mode. Table 3-2 provides the detailed results for the NCVS-1 by sample frame, mode of contact and mode of interview. Table 3-2 includes all cases for the IVR cases that were contacted by mail (IVR Mail). For modes where initial contact was conducted by telephone, only cases that completed the household screening are included. The large number of cases in the group 'other' for IVR Mail consists of cases where no mailing was returned undeliverable and contact with the IVR system was never initiated by the household. Refusals for the IVR Mail were received by a respondent that contacted a toll-free number monitored by inbound operators. Any household that called this number and refused were not sent any additional mail invitations.

Table 3-1. Detailed Frequency of Final Dispositions at the Household Screening Stage for Each Interview Mode Contacted by Telephone by Sample Frame

	ABS Frame		Police Frame	
	IVR Telephone	CATI	IVR Telephone	CATI
Interview				
Completed Household Screener	709	335	177	84
Eligible, Non-Interview				
Refusal	580	310	137	73
Language	69	42	19	9
Respondent Unavailable	123	48	28	16
Non-Contact	377	195	117	46
Other	2	1	0	0
Not Eligible				
Nonresidential	27	17	51	27
Non-Working Number	389	191	195	106
Total	*2,276*	*1,139*	*724*	*361*

Table 3-2. Detailed Frequency of Final Dispositions for the NCVS-1 for Each Mode of Contact and Interview by Sample Frame

	ABS Frame			Police Frame		
	IVR Mail	IVR Telephone	CATI	IVR Mail	IVR Telephone	CATI
Interview						
Completed Interview	1,696	375	228	642	87	50
Eligible, Non-Interview						
IVR Breakoff	274	124	–	142	26	–
Refusal	83	107	47	17	25	18
Language	–	5	5	–	1	2
Respondent Unavailable	–	78	44	–	29	10
Other	4,174	21	11	2,172	9	4
Not Eligible						
Undeliverable	232	–	–	552	–	–
Nonresidential	0	–	–	3	–	–
Total	*6,459*	*709*	*335*	*3,528*	*177*	*84*

The overall response rates for each group are shown in Table 3-3. For purposes of comparison to the CATI, the IVR cases exclude those that were offered $10. Consequently the total sample listed for the IVR cases is smaller than shown in prior tables. Since the IVR Mail group was only contacted by mail there is only a single stage of response to calculate. The two modes that initiated contact by telephone include the household screener and the NCVS 1. Response rates for the telephone contacts are calculated by multiplying the household screener response and response to the NCVS components.

Table 3-3. Response Rates by Mode of Contact and Interview*

	IVR Mail		IVR Telephone		CATI	
	Value	n	Value	n	Value	n
Overall	22.8+	3933	17.9	1,155	24.0+	1,159
Household Screener Stage	–		36.8	1,155	36.2	1,159
NCVS -1	–		48.7	425	66.3+	419

*Excludes IVR cases that were offered a promised incentive.
+Differences are statistically significant at p<.05 for IVR Mail vs. IVR Telephone; CATI vs. IVR Telephone

The overall response rate was nominally highest for the CATI group (24.0). Differences between the groups were only statistically significant for the IVR Mail and CATI groups when compared to IVR Telephone ($p<.05$).

The IVR Telephone and CATI were both contacted by a telephone interviewer who administered the household screener that identified the respondent with the next birthday. Conditional NCVS-1 level response rates were significantly different, with IVR Telephone resulting in about a 17 percentage point decrease in response. This indicates that switching from interviewer administered interview to IVR interview results in a reduction in overall response. Previous research has identified this as a concern (see. Mingay, 2000; Schneider, et al., 2005; Tourangeau, et al., 2002; Dillman, et al., 2009).

3.1.1 Sample Frame Response Rates

The response rates above combine the two different sample frames – ABS and police records. The ABS sample most closely resembles a general population sample, while the police frame includes individuals who are at greater risk of being a victim. The latter would include individuals who are from lower socio-economic groups and are more highly mobile, both of which generally respond to surveys at a lower rate. It is also the case these households had an individual who reported a crime to the police. This could either increase or decrease response to the survey, depending on whether the individual selected to participate in the survey is the same individual who reported a crime to the police.

Table 3-4 illustrates differences between each sample group (ABS and Police Department) and area (Houston, St. Louis) across the three experimental groups. This also breaks out the IVR Mail police sample between the addresses that were matched and those not matched to a telephone number. Looking first at overall response rates for each frame, the ABS frame had nominally higher response

than the matched Police frame. Within the ABS frame response was significantly higher for the IVR Mail and CATI group when compared against IVR Telephone ($p<.05$). For the Police frame no differences between contact groups was significant. While the IVR Telephone group was lowest, the sample sizes are too small to detect differences.

Table 3-4. NCVS-1 Response Rates for Each Sample Group by Mode of Contact, Interview and Metro Area

	IVR Mail				IVR Telephone		CATI	
	Value		N		Value	n	N	Value
ABS – Total	23.4		3,107		18.0	920	24.5	931
Houston	18.7		1,529		14.8	439	19.5	452
St. Louis	27.9		1,578		21.0	481	29.2	479
	Matched		Unmatched					
Police – Total	20.3	826	13.5	644	17.4	235	21.9	228
Houston	16.0	268	10.2	304	11.3	71	15.2	79
St. Louis	22.4	558	16.5	340	20.1	164	25.5	149

*Excludes IVR cases that were offered a promised incentive.

Generally, the response rates are highest in St. Louis compared to Houston. For the telephone contact modes, much of this difference is due to lower household screening rates. That is, it seems to be harder to get initial cooperation in Houston. Once on the phone, the NCVS-1 rates are very similar (data not shown). Language proficiency may be one explanation for the differences between the cities. Houston has a much higher non-English speaking population than St. Louis. According to data from the American Community Survey (ACS), 2.6 percent of the households with members age 14 or older are non-English speaking. This compares to 14.6 percent for Houston. The mailings and surveys were only fielded in English. The proportion of cases finalized as a language problem, out of those where contact with the household was made, was about 11 percent for Houston, compared to about 1 percent for St. Louis (combining sample frames).

There is a large difference between the matched and unmatched addresses for the police sample. Overall, the NCVS-1 response rate in the Police frame is 20.3 percent for the matched households compared to 13.5 percent for the unmatched households. This reflects the different types of households that tend to yield a matched telephone number. The matched households tend to be more stable and generally easier to contact by telephone. Unless otherwise stated, the response rates in the rest of this chapter exclude the unmatched addresses. This facilitates comparisons to the cases contacted by telephone, which did not include any unmatched addresses.

3.1.2 NCVS-1 Response Rates and Mode of Response Entry

The NCVS IVR instrument incorporated two modes for responding to the questionnaire (Keypad and speech). Each mode offers advantages and disadvantages. Speech response is a more natural process. Respondents do not have to alter their use of the telephone to respond. Speech is also an easier response task when the telephone keypad is located on the handset or lacks a tactile surface, such as on touchscreen smartphones. For the Keypad mode the respondent must take the telephone away from their ear to press the appropriate key for their response. The respondent risks missing part of the subsequent question while returning the telephone to their ear. However, the keypad is a more reliable method of entry. The system typically recognizes the entry, unlike speech where the system may not understand the response. In addition, when question sensitivity is a concern, respondents may be more comfortable using a keypad to enter their response, especially when there are others present at the time of the interview.

Keypad instructions for which key to press for each response option were provided after each question; these instructions could be interrupted by the respondent. Speech respondents were required to provide a voice response for each question, with the exception of questions requiring a numeric response, such as, age, frequencies of event, or dollar amounts. For numeric response questions, speech respondents were instructed to use the keypad. This was because it is difficult to capture numeric responses when they are spoken. There are a multitude of ways numeric responses can be offered and it is difficult to incorporate these into the IVR dictionary. In addition, respondents are accustomed to modifying their response, usually to imply precision, or lack thereof. For example, a dollar amount of $1,500 can be answered with "fifteen hundred", "one thousand five hundred", or "one five oh oh". Examples of modifiers that respondents' use are: "about", "around", and "dollars". These variations increase the programming complexity for a speech recognition system and increased the likelihood that a response would not be recognized.

Response rates by mode of entry for the IVR Mail and IVR Telephone groups are shown in Table 3-5 below. These addresses exclude the addresses where a $10 incentive was offered and the unmatched addresses. For the IVR Mail group, the overall response did not differ by mode of entry. For IVR Telephone respondents assigned to use the keypad had significantly higher response than those assigned to the Speech mode ($p<.05$). The differences for are at NCVS-1, where keypad mode respondents had significantly higher response.

Table 3-5. NCVS-1 Response Rates by Mode of Response for IVR Mail and IVR Telephone Groups*

	IVR Mail		IVR Telephone	
	Value	n	Value	n
Overall				
Keypad	24.3	1,954	21.0	586
Speech	21.2+	1,979	14.7	569
Household Screener Stage				
Keypad	–		38.9	586
Speech	–		34.6	569
NCVS Components				
Keypad	–		54.0x	228
Speech	–		42.6	197

*Excludes IVR cases that were: 1) offered a promised incentive and 2) did not match to a telephone number.
+Significantly different from IVR Telephone at p<.0001
XSignifiantly different from IVR Telephone speech at p<.05

This last result puts a different light on the differences between the two IVR methods of contact discussed in the last section. Focusing on just the Keypad method of entry, the differences in response rate between the IVR Mail and the IVR Telephone are greatly reduced ($p<.10$). The difference for speech is much larger ($p<.0001$). Generally, there may be more frustration with the speech entry mode because of problems with recognition (discussed later). But this may be exacerbated in the IVR Telephone setting where respondents are not necessarily expecting to be using an IVR. For the IVR Mail mode, respondents are calling in expecting to work with a computer, which lead to lower expectations with respect to the logistics of the interview.

3.1.3 Breakoffs on NCVS-1

Table 3-6 displays breakoffs for the NCVS 1 interview for the two mode groups. There were just under 17 percent of respondents who successfully accessed their assigned interview in the IVR system but did not complete the screening interview. Accessing the interview is defined as entering a valid ID and confirming the ID that was entered for the IVR Mail group, or being successfully transferred for the IVR Telephone group. Most respondents, about 44 percent, disconnected before the demographic questions early within the NCVS interview, another 28 percent disconnected before any victimization question (within the demographic section), while the remainder, about 29 percent, disconnected sometime during the victimization questions.

Table 3-6. Distribution of Breakoffs within NCVS 1 Interview by Mode*

	IVR						CATI
	Total		Keypad		Speech		
NCVS 1 – Breakoff Location	Value	n	Value	n	Value	n	
Percent Breakoffs	16.6%	559	14.2%	239	19.3%	320	18%
Location							
Before Demographics	43.5%	243	52.3%	125	36.9%	118	81%
Before Victimization Questions	27.9%	156	16.7%	40	36.3%	116	
During Victimization Questions	28.6%	160	31.0%	74	26.9%	86	19%

*Excludes IVR cases that were: 1) offered a promised incentive and 2) did not match to a telephone number.

Breakoffs are significantly higher for the speech response mode. The Keypad respondents were more likely to break off at the beginning of the NCVS-1 instrument. Breakoffs occurring at this point (before demographics) can be characterized as respondents who try a few questions, or are only contacting the system to 'see what it's like.' There is more of a tendency for the speech respondents to break off once getting into the demographic questions. This may be due to the increased level of frustration some respondents may have from the system not being able to cleanly recognize their speech.

These breakoff rates are similar to those experienced on the CATI interview, where about 18 percent broke off once starting the screening interview. It was difficult to disaggregate in the same way as the IVR, but it is apparent that many of these respondents are dropping before the victimization questions (81 percent).

3.2 NCVS-1 Respondent Characteristics

Distributions of selected household and respondent characteristics were examined for each group. This review provides insight into whether the different methods of contact and interview administration appeal to different population groups. The characteristics in the tables below were collected for the person selected to complete the NCVS interview. With the exception of household income, all characteristics below were collected early in the NCVS 1 interview, prior to questions about victimization. Income was asked as one of the last items in the interview and after the debriefing questions. For the CATI group, all characteristics were collected through the CATI interview; for IVR Mail these were collected in the IVR interview; and for IVR Telephone, these were collected in the IVR interview, with the exception of number of adults living in the household which was collected within the CATI screener.

Tables for all variables are produced for the ABS sample frames combining across Houston and St. Louis sample areas. Small samples sizes for the police frames make it difficult to produce meaningful conclusions. The ABS sample frame would most closely represent a general population survey, although the sample is for addresses that match to a telephone number. These households will tend to represent less mobile, older individuals. As in previous sections, addresses where an incentive was offered or where there was no match to a telephone number are excluded.

The total number of persons and number of adults are provided in Table 3-7 below. The differences are not large. But it is the case that respondents completing interviews in the IVR Mail group were somewhat smaller than the IVR Telephone and CATI group ($\chi^2 < 0.05$).

Table 3-7. Household Size for Each Sample Group by Mode of Contact and Interview*

	IVR Mail		IVR Telephone		CATI	
	Value	n	Value	n	Value	n
# Persons in Household[1]						
One	25.4	185	29.1	48	31.8	71
Two	36.8	268	36.4	60	39.0	87
Three to Five	35.5	258	24.2	40	27.8	62
Six or More	2.3	17	10.3	17	1.3	3
# Adults in Household[2]						
One	29.0	210	35.8	59	37.4	85
Two	51.0	369	44.2	73	52.0	118
Three to Five	19.8	143	16.4	27	10.6	24
Six or More	0.3	2	3.6	6	–	-

*Excludes IVR cases that were: 1) offered a promised incentive and 2) did not match to a telephone number.
Difference between IVR Mail, IVR Telephone and CATI-Only are different at p<.05
1 Table does not show respondents reporting 0 (1 respondent from IVR Telephone). Also excludes 5 missing values from CATI respondents.
2 Table does not show respondents reporting 0 (4 respondents from IVR Mail and 1 respondents from IVR Telephone). Also excludes 1 missing value from CATI respondents.

Tables 3-8 and 3-9 provide selected demographic characteristics of the respondent across three groups. . For age the IVR Telephone group included the largest proportions of respondents 65 and older (p<.05).[2] IVR Mail and CATI differ a bit in the younger age groups, although the total number of respondents in these groups is quite small. Marital status and gender were not significantly different. There were significant differences for race and ethnicity. . The proportion identifying as Hispanic was highest for IVR groups (IVR Mail and IVR Telephone), although not by a large percentage (7 percent vs. 4.5 percent). The proportion identifying as two or more races is the

[2] Due to the low number of respondents reporting age 18 to 24 this group was combined with the 25 to 34 category for conducting the chi-square test. For race the low numbers of respondents identifying as American Indian or Alaskan Native resulted in this group combined with the Asian or Pacific Islander category for conducting the chi-square test.

reason for the difference in race. This is an anomaly with the IVR system. When reviewing these cases, it was found the IVR system mis-recognized a come cases and falsely attributed a second race to these individuals. For respondent education, proportions were not statistically different across the three groups.

Table 3-8. Respondent Characteristics for Each Sample Group by Mode of Contact and Interview*

	IVR Mail		IVR Telephone		CATI[1]	
	Value	n	Value	n	Value	n
Age						
18 to 24	1.8	13	1.2	2	2.8	6
25 to 34	6.0	44	4.2	7	1.4	3
35 to 49	22.7	165	19.3	32	21.1	46
50 to 64	37.8	275	29.5	49	37.2	81
65 and Older	31.7	231	45.8	76	37.6	82
Marital Status						
Married	58.4	425	57.2	95	53.5	121
Widowed	11.0	80	16.9	28	15.0	34
Divorced	15.8	115	14.5	24	14.2	32
Separated	1.5	11	15.4	2	1.8	4
Never Married	13.3	97	10.2	17	15.5	35
Gender						
Male	44.4	323	44.0	73	40.8	93
Female	55.6	405	56.0	93	59.2	135
Race & Ethnicity[2]						
Hispanic	7.0	51	7.8	13	4.5	10
White	71.8	523	62.7	105	71.4	160
Black	11.3	82	6.6	11	16.1	36
American Indian or Alaskan Native	0.4	3	–	-	0.4	1
Asian or Pacific Islander	4.3	31	3.0	5	4.9	11
Two or More Races	5.0	36	17.5	29	2.2	5
Some Other Race	N/A	N/A	N/A	N/A	0.4	1

* Excludes IVR cases that were: 1) offered a promised incentive and 2) did not match to a telephone number.
1 Excludes CATI respondents reporting refused or don't know.
2 Due to the structure of the IVR interview, respondents could actively not identify a race. For IVR Mail 0.3% of respondents did not identify a race, for IVR Telephone this was 2.4%.

Table 3-9. Respondent Education for Each Sample Group by Mode of Contact and Interview*

	IVR Mail		IVR Telephone		CATI[1]	
	Value	n	Value	n	Value	n
Less than High School	4.5	33	7.2	12	6.7	15
High School Graduate or GED	27.8	202	29.5	49	21.8	49
Some College/AA Degree	18.4	134	24.7	41	24.4	55
4-Year Degree/BA or BS	23.6	172	16.9	28	23.6	53
Attended Graduate School – No Degree	4.5	33	5.4	9	2.7	6
Graduate or Professional Degree	21.2	154	16.3	27	20.9	47

*Excludes IVR cases that were: 1) offered a promised incentive and 2) did not match to a telephone number.
1 Excludes CATI respondents reporting refused or don't know.

Table 3-10 shows housing characteristics across the three groups. The two IVR groups have some individuals who did not report a residence type at all. There is a much larger proportion for the IVR Telephone group. This might have occurred from converting these questions to the yes/no format. Respondents were asked about each residence type separately in IVR and could say no to all options. This would result in an unknown housing type. If these cases are removed, there are similar proportions across the three contact groups (IVR Mail: 88.2 percent; IVR Telephone 84.8 percent; CATI 84.5 percent). Years at current residence significantly differed among the three contact groups ($p < 0.05$). The IVR Mail and CATI groups have similar distributions, while the IVR Telephone group had a larger proportion of respondents who had lived at their current residence for a shorter period of time.

Table 3-10. Housing Characteristics for Each Sample Group by Mode of Contact and Interview*

	IVR Mail		IVR Telephone		CATI[1]	
	Value	n	Value	n	Value	n
Residence type						
Unknown	3.4	25	12.7	21	N/A	N/A
Owned or Being Bought	85.2	620	74.1	123	84.5	191
Rented	10.9	79	12.7	21	13.3	30
Some other Arrangement	0.6	4	0.6	1	2.2	5
Years at Residence						
Less than One Year	5.0	36	3.0	5	0.9	2
1 to 5	18.5	135	31.3	52	19.3	44
6 to 10	19.6	143	12.7	21	20.2	46
11 to 25	34.9	254	27.1	45	36.4	83
More than 25 Years	22.0	160	25.9	43	23.3	53

Table 3-10. Housing Characteristics for Each Sample Group by Mode of Contact and Interview* (Continued)

	IVR Mail		IVR Telephone		CATI[1]	
	Value	n	Value	n	Value	n
Times Moved (in last 5 yrs.)[2]						
Zero	81.0	590	80.7	134	85.1	194
One	9.9	72	8.4	14	10.5	24
Two	5.0	36	7.2	12	3.5	8
Three or More	4.1	30	3.6	6	0.9	2

*Excludes IVR cases that were: 1) offered a promised incentive and 2) did not match to a telephone number.
1 Excludes CATI respondents reporting refused or don't know.
2 For CATI group, this question was asked only for respondents who lived at current address less than 5 years. For IVR Mail and IVR Telephone this was asked for respondents who lived at current address less than 6 years.

While distributions for years at current residence were significantly different, this was not the case for number of times moved within the last five years. Nominally the CATI group had more respondents who have not moved within the last five years, and the IVR groups had more high frequency movers.

Table 3-11 provides the distributions for income. This was asked at the end of the questionnaire. Twenty-two percent of the IVR respondents broke off either before or at this item. Approximately 7 percent broke off when getting to the income question. The remaining 15 percent broke off after completing NCVS-1, but before the income question. Many of these additional breakoffs were respondents not completing their pre-requisite number of NCVS-2 modules. For CATI, no one broke off at or before the income question.

Table 3-11. Household Income for Each Sample Group by Mode of Contact and Interview[1]*

	IVR Mail		IVR Telephone		CATI	
	Value	N	Value	n	Value	n
Less than $1,000	16.4	81	17.8	18	–[2]	–[2]
$1,000 to $7,499	6.3	31	15.8	16	5.2	10
$7,500 to $14,999	4.0	20	5.9	6	5.2	10
$15,000 to $24,999	5.3	26	5.0	5	4.1	8
$25,000 to $34,999	6.7	33	5.0	5	10.3	20
$35,000 to $49,999	7.9	39	4.0	4	11.3	22
$50,000 to $74,999	16.6	82	15.8	18	21.6	42
$75,000 or more	36.8	182	30.7	31	42.3	82

*Excludes IVR cases that were: 1) offered a promised incentive and 2) did not match to a telephone number.
1 Excludes 234 (32.1%) of respondents for IVR Mail reporting zero or missing, for IVR Telephone this excludes 65 (39.2%) respondents. Excludes CATI respondents reporting refused or don't know.
2 Not available as an option for CATI group, respondents to this category would be considered $7,499 or less in CATI This category is a combination of the categories 'less than $5,000' and '$5,000 and $7,499'.

The placement of the income question differed slightly between the IVR and CATI. For IVR groups this was the second to last question and followed debriefing questions on perceptions of police and perceptions of the IVR system. For CATI it was asked before perceptions about police.

The main income question was identical across the IVR and CATI. Respondents were asked for the total combined income of all household members. The IVR respondents were asked to enter their responses using the keypad, even if they were in the Speech entry group. The IVR instrument specified whole dollars since respondents were required to enter this information, while in CATI respondents reported this number to an interviewer that selected the appropriate income category. The IVR instrument accepted any numeric response that was seven digits or less. Entries of 400,000 or more were verified. The CATI interviewer, however, could verify any response to ensure the respondent is providing the total combined income of all household members for the appropriate time period.

On the IVR, a number of respondents reported '0' as a way to not answer the question. This answer was accepted as a legitimate response. A total of 34 respondents entered this as a response (12.1 percent). On the CATI, if the respondent reported "refused" or "don't know" the interviewer followed with less specific questions asking for household income. These would ask if the value was more or less than an anchor point of. The unfolding process would continue until the next one of the income categories was reached. Out of the CATI respondents that said refused or don't know to the initial income question, 29.2 percent (14 cases) provided a response to the unfolding set of income items.

There are significant differences between the IVR and CATI groups. For the two IVR groups, the distributions are relatively similar as illustrated in Table 3-11. But there are very large proportions in the lowest income category of 'less than $1,000'. For the CATI group, there is a much smaller percentage of those falling in the lowest income group.

We believe that these differences are due to entry error on the part of the IVR respondents. The extremely low income groups may be the result of entering data for the incorrect units, such as individual income (not family) or for a non-annual time unit (e.g., monthly or weekly). A second possibility is not entering enough digits. Respondents may lose track of how many times they have pressed '0', or on touchscreen telephones, the touch may have been too brief or rapid to detect all selections for '0'.

To get an idea of how the composition of the respondents compare to the general population, Table 3-12 compares selected characteristics with the distributions for these areas taken from the American Community Survey (ACS). The ACS provides information on the non-institutional population, which is different from the matched-address NCVS IVR sample. Nonetheless, this provides some perspective on the profile of the respondents to the survey. In particular the IVR sample severely underrepresents the young, lower educated and Hispanics. This result is similar for all of the modes of interviewing and contact.

Table 3-12. Comparison of ACS and NCVS IVR ABS respondent characteristics*

		ABS Sample*		
	ACS	IVR Mail	IVR - Tel	Interviewer
Single person household	25.6	25.4	29.1	31.8
Females	51.3	55.6	59.2	56.0
Age 18 to 34	32.7	7.8	4.2	5.4
Hispanic	21.9	7.0	4.5	7.8
Non-Hispanic Non-White	23.4	21.0	24.0	27.1
Education is HS or less	16.9	4.5	6.7	7.2

*Excludes IVR cases that were: 1) offered a promised incentive and 2) did not match to a telephone number.

3.3 Non-Response on NCVS-2

The above analysis describes response to NCVS-1. One of the concerns with an IVR application for the NCVS is whether respondents will complete not only the screener, but also the detailed incident form (NCVS-2). Table 3-13 provides the response rates for each of the two sample frames (ABS and Police) for completing NCVS-1 and for completing both NCVS-1 and NCVS-2. This rate is what is typically reported for the ongoing NCVS because it includes both NCVS-1 and NCVS-2. For the ABS sample, the IVR Mail rates drop by several percentage points (e.g., 23.4 vs. 21.4 percent for IVR-Mail).

Table 3-13. Comparison of total response rates by Mode of Contact and Mode of Interview+

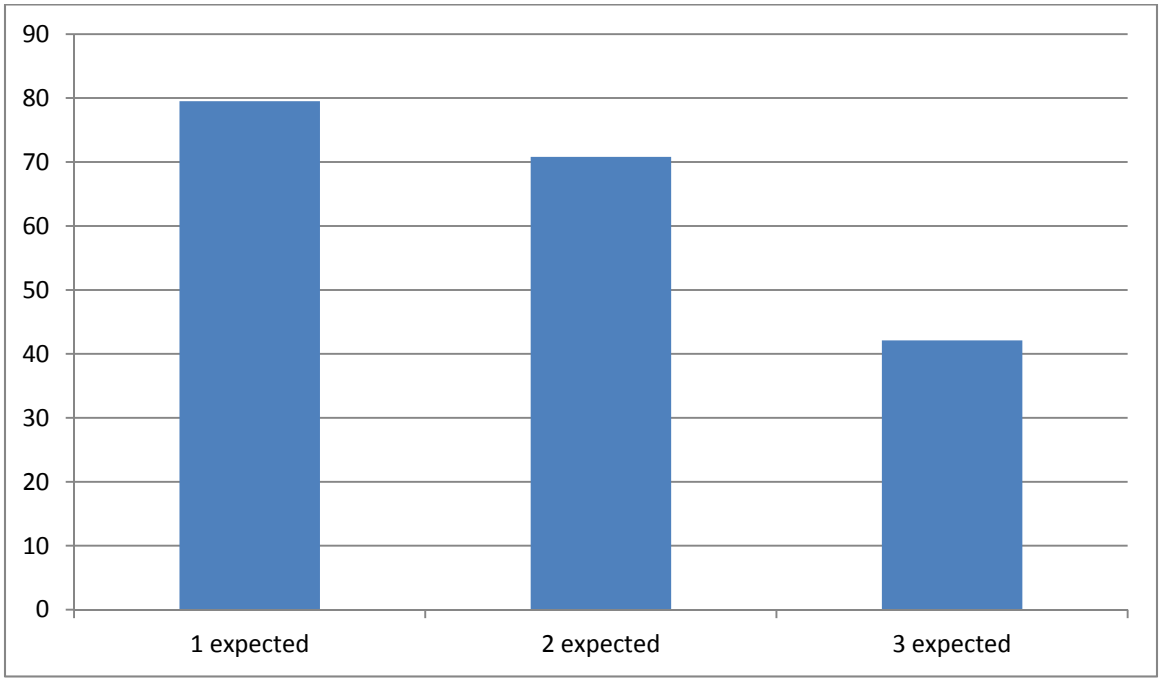
	IVR Mail		IVR Telephone		CATI	
	NCVS-1	NCVS-1 and 2	NCVS-1	NCVS-1 and 2	NCVS-1	NCVS-1 and 2
ABS	23.4*	21.4	18.0	16.6	24.5	24.5
n	3,107	3,107	920	920	931	931
Police	20.3**	14	17.4*	11.9	21.9	21.9
n	826	826	235	235	228	228

+Excludes IVR cases that were: 1) offered a promised incentive and 2) did not match to a telephone number.
*Difference between NCVS-1 vs. NCVS1&2 is significant at p<.10
** Difference between NCVS-1 vs NCVS1&2 is significant at p<.01

The largest drop in rates is for the Police sample. The IVR Mail drops from 20.3 to 14.0 and the IVR Telephone drops from 17.4 to 11.9. In comparison to the IVR, none of the telephone respondents dropped out between NCVS-1 and NCVS2. This holds for both the ABS and the Police sample

Overall, approximately 30 percent of respondents did not complete all of the required NCVS-2 incident forms. Completing all of the forms is related to the number of incident forms that are being requested. This is illustrated in Figure 3-1, which provides the percentage of respondents that completed all of the expected NCVS-2 forms by the number expected. The rate is about 80 percent for those that were expected to complete one NCVS-2. This drops to around 40 percent for those expected to fill out three forms.

Figure 3-1. Percent of Persons completing all NCVS-2 forms by the number of forms expected*



*Excludes IVR cases that were: 1) offered a promised incentive and 2) did not match to a telephone number.

As noted above, there was a significant drop-off in response for NCVS-1 for speech respondents (Table 3-14). When looking at the percentage of respondents that complete all of the forms by mode of entry, there is a statistically significant difference between the percentage of respondents that fill out all of their respective incident forms ($p<.001$). Among those that had speech entry, 73.3 percent completed all forms, compared to 80.7 percent for those that used the keypad.

Table 3-14. Percent of Missing NCVS-2 forms by Mode of entry*

	Mode of Entry*	
# of missing NCVS-2	Speech	Keypad
0	73.3	80.7
1	15.4	12.1
2	7.4	5.7
3	3.9	1.5
Total	100.0	100.0
N	434	477

*Excludes IVR cases that were: 1) offered a promised incentive and 2) did not match to a telephone number.
Difference between incentive and no incentive is significant at p<.001

3.4 Summary and Implications

This analysis provided evidence that NCVS-1 response rates for the IVR Mail are equivalent to CATI. The IVR Telephone methodology had the lowest response rates. This was largely because there were a significant number of individuals who abandoned the interview during the transfer between the interviewer and the IVR. The largest dropoff occurred for the respondents using speech as the mode of entry. This dropoff has been noted in other IVR surveys using a similar methodology (e.g., Gribble, et al, 2000). With respect to respondent characteristics, the IVR Mail and the CATI were the most similar, with the IVR-Telephone having more older respondents, individuals with a marital status of 'separated' and fewer people who owned their home. There were very large differences between all of the experimental groups and the general population. When comparing to ACS data, all of the groups underrepresented young people, those with a lower education and Hispanics.

The IVR response rates drop once considering completion of NCVS-2. Overall, the IVR-Mail and IVR-Telephone rates for the ABS sample drop 2 – 3 percentage points. While this is a relatively small drop in the overall rate, this hides the fact that a significant number of individuals did not complete all of the NCVS-2 forms that were expected. Overall, approximately 30 percent of the respondents did not fill out all of the required incident forms. This compares to a 100 percent completion rate for the CATI.

At least from a response rate perspective, these results suggest that it is feasible to screen large numbers of individuals using an IVR Mail procedure. If the options are a CATI or IVR Mail, the two seem very equivalent from a response rate perspective for NCVS-1. The methods were also equivalent with respect to the demographic composition of respondents who completed the NCVS-

1. Using the IVR to also complete NCVS-2 is also possible. A significant number of respondents did complete all of the expected forms. However, there was a significant drop-off in response, with 30 percent of respondents not completing all of the forms required of them.

These results do not directly address the use of IVR as a way to interview core NCVS respondents, for example at second or later panel waves. When compared to the CATI interview, however, there was no evidence that respondents were less willing to complete the NCVS-1, once they started it. The percentage of individuals who broke off once starting the screener was virtually the same between the IVR and the CATI. In fact, this percentage was lower for the Keypad respondents. This supports the general idea that if respondents understand they are going to be using a computer to take a survey, they are not necessarily adverse to completing the task, at least when the task is relatively short, such as the NCVS-1. As noted above, completing the NCVS-2 is more problematic. However, in the context of the NCVS panel design, there is an opportunity to follow-up respondents who do not complete all NCVS-2 forms. NCVS-2 respondents may also be more motivated to complete the entire survey than the respondents included in the current study.

Encouraging Response & Enhancements to Increase Response

4

Chapter Highlights and Key Findings

- The promised incentive had a significant effect for the IVR Mail mode, raising the overall response rates by 6 - 7 percentage points depending on the sample frame. It had an inconsistent effect for the IVR Telephone mode of 0 – 4 percentage points.
- For the IVR Mail, the insert significantly raised the response rate when used in conjunction with the promised incentive. In the incentive condition, it raised the response rate approximately 3 additional percentage points. When comparing the incentive-with-insert to the no-incentive conditions, the difference in response rate is approximately 10 percentage points. The insert had no effect when an incentive was not offered.
- There were no meaningful differences in the demographic composition between the different incentive/insert groups.
- There were indications that the incentive increased motivation for respondents to complete the survey. Respondents receiving the incentive were significantly more likely to complete all of their NCVS-2 forms and had less missing data for the income.
- Take-away point: A small promised incentives is effective for the IVR Mail group. If an IVR procedure is used, either at a local area level or as part of the ongoing NCVS, an incentive will significantly increase response rates and the overall efficiency of the methodology.

There were two experiments implemented to increase response rates for the IVR. In one experiment half of the IVR addresses were randomly assigned to a group that was promised a $10 incentive upon completion of the survey. The IVR Mail group was informed of the promised incentive through wording in the invitation letter. Respondents would also be reminded of this at an early point during the IVR interview. Mention of the promised incentive was made in each follow-up invitation letter that was mailed. The IVR Telephone group also received wording in the prenotification letter and were told by the telephone interviewer just prior to transfer to the IVR system.

The other experiment manipulated the inclusion of an insert within the mail invitation letter for the IVR Mail group. Half of the addresses were randomly assigned to receive an insert with their invitation letter. This experiment was fully crossed with the incentive experiment, which affected the text of the insert. Addresses assigned to the insert condition and the incentive condition had

wording on the insert that they would receive $10 for completing the interview. Those assigned to the no incentive group were sent an insert that had wording to appeal for their help by completing the interview. The insert also included the number for contacting the IVR system and their unique ID assigned to the address.

The results of each experiment are presented below by looking at response rates for each group. Since the two experiments were fully crossed we also examine interaction effects.

4.1 Incentive

Table 4-1 shows the response rates for the IVR Mail and the IVR Telephone groups for the incentive conditions. This table excludes cases assigned to the insert condition as that experiment was only implemented for the IVR Mail group. This also excludes the unmatched addresses sent to the police sample. Looking at the IVR Mail group, there is a significant effect of the incentive for both the ABS and Police frame ($p<.01$). There is a smaller effect for the IVR Telephone group. The effect for the ABS sample is significant ($p<.05$), but not significant for the Police Frame.

Table 4-1. Response Rates by Incentive Condition for IVR Mail and IVR Telephone Groups by Sample Frame (sample sizes in parentheses)*

	IVR Mail		IVR Telephone	
	Value	n	Value	n
ABS				
Incentive	29.3+	1,566	22.1x	940
No Incentive	23.1	1,548	18.0	920
Police				
Incentive	24.5 +	739	18.9	243
No Incentive	17.3	728	17.4	235

*Excludes IVR cases that were: 1) had an insert and 2) did not match to a telephone number.
+ Difference between incentive conditions is significant at p<.01;
X Difference between incentive conditions is significant at p<.05

The smaller significant effect in the IVR Telephone could be attributable to several factors. One is that promised incentives are difficult to communicate over the telephone. Respondents may not have been aware of the promise or discounted it because they did not believe the offer. A related explanation is that the IVR Mail group may have been more likely to read and notice the offer of $10. The IVR Mail group had to use the letters to call in. The letters mentioned the promised incentive. The respondent to the IVR Telephone may not have seen the letter, either because someone else in the household opened it or the letter never arrived at the house because the address

did not match the phone number (about 6 percent of the cases). The first time they heard about the incentive was during the telephone interview.

When disaggregating the Police data by city, there is a stronger effect of the incentive for the IVR Telephone for St. Louis when compared to Houston (data not shown). This may explain some of the null effect for the Police data. The data provided in the previous chapter indicated there are more households in Houston with individuals do not speak fluent English. It may be that there were more respondents in Houston who were not as likely to understand the interviewer with respect to the incentive.

4.2 Mail Insert

The insert was only administered to the IVR Mail group, and was only included in the initial mailing. Subsequent or follow-up mailings did not include an insert. The wording on the insert was different based on assignment to the incentive condition. Table 4-2 provides response rates for the IVR Mail group crossing the incentive and insert experiments. Within the no incentive group, the insert and no insert groups had nearly identical rates of response. For the addresses assigned to the incentive condition, the insert resulted in significantly higher response than the no insert condition ($p < 0.05$). The effect of the incentive is significant within each insert experimental/ condition. Addresses assigned to a promised incentive offer had significantly higher response across both insert conditions (Insert – $p < 0.0001$; No Insert – $p < 0.0001$). This effect is larger for those assigned to also receive the insert. The message offering a promised incentive on the insert appears to have increased the saliency of the incentive offer.

Table 4-2. Response Rates by Incentive and Insert for IVR Mail Group

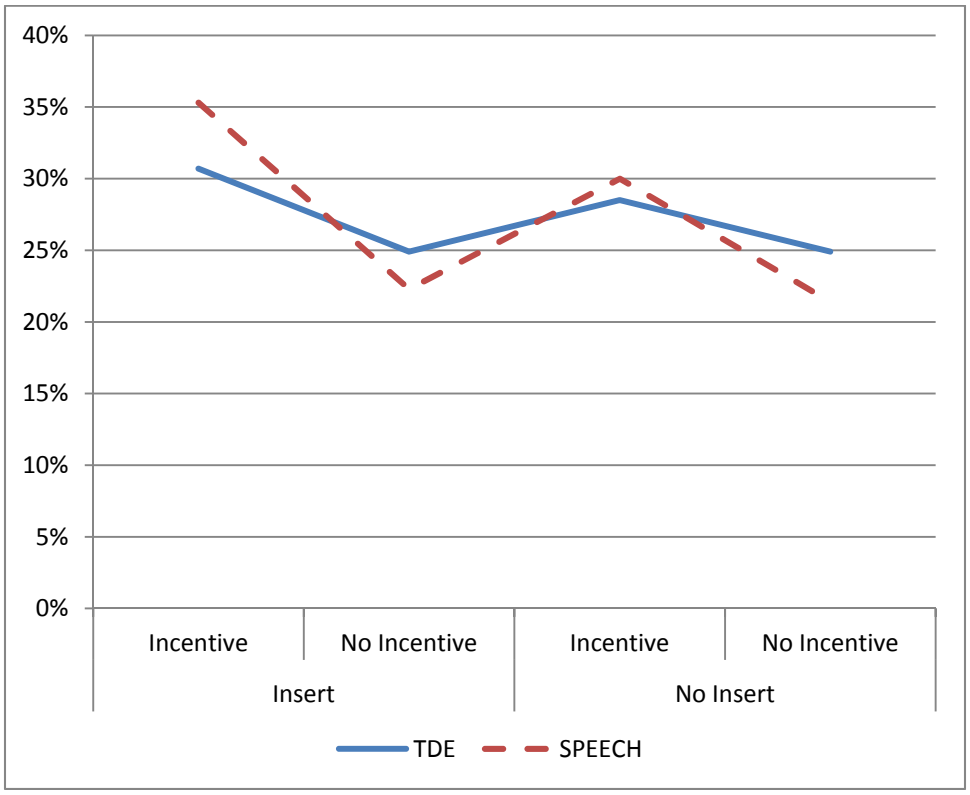
	Insert		No Insert	
	Value	n	Value	n
Incentive	32.9*+	1554	29.2+	1566
No Incentive	23.6	1559	23.1	1548

*Significant difference between the insert and no insert condition at p<.05.
+ Significant different different between the incentive and no incentive condition at p<.0001

Taking a final look at the incentive and insert experiments by mode of response, Figure 4-1 provides the response rates by incentive and insert groups by each mode of response. While the differences are generally small, the incentive appears to exert a greater impact on Speech mode respondents. As will be discussed below, the speech mode exhibited more signs of user-problems. The larger effect

of the incentive for the speech group may be an indication that the incentive provided an additional motivating factor for respondents to finish, once they started the survey.

Figure 4-1. Response Rates by Incentive and Insert Experimental Groups, for Keypad and Speech Response Modes (ABS Sample Frame)*



* Interaction between mode of entry and incentive is significant at p<.05. Excludes addresses that did not match to a telephone number

4.3 Incentive and Demographic Composition

One reason to enhance response rates is to try to encourage groups that might be under-represented in the survey. An analysis was conducted that compared demographic characteristics across the different incentive – insert groups: 1) promised incentive with insert, 2) promised incentive only, 3) Insert only and 4) No incentive or insert. This analysis was completed with just the IVR Mail group, since this was the only one that exhibited a large and consistent effect of the promised incentive. Overall, there were no significant differences among the different treatment groups. There was some indication that the no incentive conditions resulted in more single person households. There was also some indication that the incentive brought in more black respondents. The most notable difference was that the incentive group had less missing data for the income item. Among those

provided an incentive, about 24 percent did not provide their income in comparison to around 32 percent for those that did not receive an incentive.

4.4 Incentive and the Detailed Incident Form

As noted in Chapter 3, there was a significant drop-off in response for NCVS-2. One possible benefit of an incentive is that it increases the number of respondents that finish the entire survey, including any NCVS-2 forms. When looking at the percentage of respondents that complete all of the forms (Table 4-3), there is a statistically significant effect of the incentive ($p<.001$). Among those that received an incentive, 73.5 percent completed all forms, compared to 61.6 percent for those that did not receive an incentive.

Table 4-3. Percent of Missing NCVS-2 forms by Incentive Condition

	Incentive*	
# of missing NCVS-2	No	Yes
0	61.6	73.5
1	18.4	15.9
2	13.6	6.3
3	6.4	4.3
Total	100.0	100.0
N	375.0	536.0

*Difference between incentive and no incentive is significant at p<.001

4.5 Summary and Implications

The results in this chapter support the use of a promised incentive of $10 to encourage use for the IVR. When paired with the insert, the response rate went up by approximately 10 percentage points. The promised incentive did not work as well for the IVR Telephone group. There are several possible explanations for this. The first involves the mode of contact. The IVR Mail group had to receive the mail invitation and actually read the material in order to access the IVR survey. In addition, this group all received the $2 incentive included in the pre-notification letter. Both of these conditions increases the likelihood the respondent will notice the offer of the incentive. IVR Telephone respondents were also sent a prenotification letter with $2, but they may not have read the letter. The telephone number may not have matched the address to which the letter was mailed.[3]

[3] About 6% of the telephone respondents reported their address was not the same as the one the letter was sent to.

Or the person that opened the letter may not have been the telephone respondent. Receipt of $2 prior to participation may increase the saliency, and the credibility of the offer of the promised incentive. These results are in line with past research involving promised incentives for interviewer administered surveys by telephone (Cantor, et al., 2008). The promised incentive may not have been salient enough to overcome the negative impact on response resulting from the transfer from the CATI to the IVR system. It also may not have been viewed as a credible offer.

This evidence reinforces the results in Chapter 3 when comparing to CATI. The ability to enhance the response rate with a small incentive is encouraging, both from the point of view of doing local area surveys and asking NCVS respondents to call in to complete the NCVS-1. The incentive also showed some evidence that data quality was enhanced with more respondents filling out all detailed incident forms and less missing data on income. However, there was no evidence that the enhanced response rate significantly improved the types of demographic groups that were represented among respondents.

Usability 5

Chapter Highlights and Key Findings

- The average time to complete the NCVS-1 ranged from 7 to 10 minutes, depending on if victimization is reported. The overall average time to complete, including filling out NCVS-2, was around 11 minutes.
- The average time to complete with one victimization was 16 minutes, 22 minutes for two victimizations and 24 minutes for three victimizations.
- Overall 62 percent of respondents encountered some type of system error on NCVS-1. This was much greater for the speech respondents (87 percent) compared to the keypad respondents (38 percent). The main reason for the high rate for the speech respondents was the inability of the system to recognize the responses.
- Respondent perceptions about the interview varied by whether a victimization form was filled out.
- Several types of errors were found with the design of the IVR. Income and selected 'Mark all that apply items' had inaccurate data. The IVR system also had isolated instances where data were miscoded.
- Take-away point #1: The time to complete the NCVS-1 is within range of many IVR surveys. Consistent with the response rate analysis, completing this portion of the interview with IVR does not seem to be overly burdensome to respondents.
- Take-away point #2 : The time to complete the NCVS-2 is significantly longer than NCVS-1. Using a keypad entry and an incentive will reduce dropping out to some extent.
- Take-away point #3. Further reductions in dropout rates will require other adaptations of the NCVS-2 to IVR. This would include reducing the amount of information collected and/or using more verbatim response options.
- Take-away point #4 : It is recommended to use a keypad entry mode as the primary method of entry, perhaps giving an option to use speech if the respondent desires. Primary reliance on speech should be considered as the speech recognition technology evolves and reduces the overall error rates.

This is the first of three chapters that address the question of whether the NCVS can be adapted for an IVR mode of interviewing. One overall measure of the success of adapting the interview is whether respondents could get through the survey and the extent they had problems during the

process. This chapter presents data on four measures of usability: 1) Interview time, 2) Measures of error of navigating and entering data, 3) Respondent perceptions about the interview experience, and 4) Other problems identified.

5.1 Length of the Interview

The mean times for specific sections are shown in Table 5-1. Timings are provided for all cases, and separately for each response mode. Extreme values for each interval were excluded from calculations for each mean. These generally resulted from instances where the respondent disconnected from the system (or in the case of errors was disconnected by the system) and called back another time to continue the interview. Excluding these, and cases for which timing data was not available, resulted in 2,625 full and partially completed interviews (at least NCVS 1 completed). This excludes the addresses that did not match to a telephone number.

Table 5-1. Average time to complete subsections of the IVR by Mode of entry

	All Cases	Keypad	Speech Only
Demographics	2.67 (2,620)	2.53 (1,384)	2.83 (1,236)
NCVS-1	6.14 (2,623)	5.94 (1,384)	6.37 (1,239)
First NCVS-2	6.70 (790)	6.30 (455)	7.24 (335)
Second NCVS-2	5.30 (300)	5.00 (173)	5.70 (127)
Third NCVS-2	5.09 (109)	4.80 (68)	5.57 (41)
Debrief with NCVS-2	3.96 (594)	3.74 (346)	4.27 (248)
Debrief No NCVS-2	3.96 (1,363)	3.76 (726)	4.18 (637)

*sample sizes in parentheses. Excludes addresses that did not match to a telephone number.

The demographic section took approximately 2.7 minutes. The crime screening questions took an additional six minutes. Examining each response mode, the speech mode took slightly longer to complete across both sections within NCVS 1. For the NCVS-2, the same pattern is apparent as was observed in NCVS 1. The speech mode was consistently longer than the keypad. Looking across the three intervals related to the NCVS-2, the timing for each subsequent victimization report

decreases.[4] Respondents, having been exposed to the questions within NCVS 2, learned what to expect and become more fluent in navigating the system. They may also become a bit more impatient and tried to work through the questions faster.

Overall, the NCVS IVR interview was relatively short victimizations (Table 5-2). The average time for this group across all eligible interviews was a little over 11 minutes. For respondents reporting at least one, two and three victimizations, the average times are 16, 22 and 24 minutes, respectively.

Table 5-2. Average time to complete NCVS sections by Victimization Status and Mode of entry

Component	Time (minutes)	Speech	Keypad
NCVS-1	7.74	8.12	7.39
No victimization	1,558	745	813
NCVS-1	10.4	10.85	10.01
Victimization reported	1,060	490	570
NCVS-1 & NCVS-2	16.46	17.72	15.53
One victimization completed	488	206	282
NCVS-1 & NCVS-2	22.47	23.44	21.69
Two victimizations completed	190	85	105
NCVS-1 & NCVS-2	24.52	31.02	28.62
Three victimizations completed	109	41	68
NCVS-1 or NCVS-1 & NCVS-2	11.65	11.91	11.41
Across all interviews	2,617	1,234	1,383

Comparing keypad response to speech response the speech mode was consistently longer across all intervals. This result is somewhat surprising. An advantage of the speech response mode is that the keypad does not need to be utilized, and instructions on how to respond are generally not needed. Keypad response can be more difficult on telephones where the keypad is on the handset and the respondent is either required to use speakerphone or remove the handset from their ear after each response. With touchpad cellphones the lack of a tactile surface may make keypad response more prone to error. We suspect that the longer length for the speech is related to the larger number of errors made when using speech. We discuss this in the next section.

[4] In some cases, fewer questions are asked after the first NCVS-2 form is filled out. If it is from a screener item that is different from the prior NCVS-2, the respondent is asked to confirm the victimization and the frequency for which it occurred within the last 12 months. If the report is an additional incident within the same screener item, these questions were not asked.

5.2 Problems Entering Responses

In addition to the timings to complete the IVR, the study collected indicators of problems with entering responses. For the keypad mode, there are timeout and invalid entry errors. Timeouts occur when the respondent does not provide an entry within a specified period of time. For most questions respondents are given six seconds to provide a response. Some questions, which may require more thought, are given 10 seconds. When timeouts occur the IVR system will play a prompt informing the respondent that it did not receive a response, provide a reminder of how to access help, and replay the question.

Another type of error for keypad mode respondents is invalid entries. Invalid entries are keypad entries that are not accepted by the IVR. For example, pressing '3' for a yes/no question where '1' is yes and '2' is no. When this occurs the respondent may not be aware of the error. The IVR system will play a prompt telling the respondent they provided an invalid response, what was entered, then replay the question.

Similar errors can occur for speech mode respondents, but for slightly different reasons. As with keypad respondents timeouts can occur when the respondent does not provide a speech response. However, these can also occur in cases where the speech response is too low to be detected by the speech recognition system. When this occurs a prompt is played that informs the respondent the system did not hear them and will repeat the question.

Invalid entries in the speech response mode differ from keypad mode responses. As with keypad responses invalid entries may be due to respondents providing an unexpected entry. For example, due to not providing an acceptable response, such as, saying "Black" in response to a question asking if they are Black or African American. In this example the expected response is yes or no. Invalid entries may also be unrecognized responses due to surrounding noises, additional speech disfluencies, or an inability of the system to determine with a degree of confidence what the speech utterance was. When this occurs, the IVR system plays a prompt that informs the respondent it did not understand them, reminds them of how to get help, and replays the question. In the case of yes or no questions, the system will also remind the respondent to say yes or no.

Errors of the types described above could occur at any question. If a respondent encountered four consecutive errors at a particular question, the system would tell the respondent it was having difficulty and end the interview asking the respondent to call back another time. This was done to

force respondents to take a break and reduce the potential buildup of frustration. While reasons for the difficulty resulting in errors is unknown, they could be attributed to the respondents location (ambient noise), situation resulting in impatience and rushing through the interview, or telephone line quality.

Table 5-3 provides data on the proportion of respondents that had none, as least one, or multiple errors as described above within NCVS 1. Looking across all cases, 62 percent encountered at least one error, with 38 percent never having any type of error. When these are separated by whether the case is a full complete or when the respondent did not complete all NCVS-2 forms (partial complete), the proportion of errors between the two groups differ. The proportion of cases with at least one error is 60 percent for completed cases, and roughly 16 points higher at 76.3 percent for partial completes. The difference for proportion of cases with five or more errors is equally as large (9.6 percent compared to 23.4 percent).

Table 5-3. Proportion of cases experiencing errors (time out / invalid response) in NCVS 1 by completion status

	None	1 or more	5 or more	10 or more
All Cases[1]	38.0%	62.0%	11.3%	1.1%
Completes Only	40.0%	60.0%	9.6%	0.9%
Partial Completes Only	23.7%	76.3%	23.4%	2.3%

[1] All cases included all completed IVR Mail and IVR Telephone including partial completes (n = 2,800). Completes only include 2,454 cases. Partial completes only include 346 cases which completed the NCVS-1 but did not complete all of the detailed incident forms..

Table 5-4 provides these data broken out by response mode. Looking at each mode of response separately, increases in proportion of cases encountering at least one error are higher for partially completed cases. Comparing keypad entry and speech response, the speech mode consistently shows higher proportions of cases encountering errors. The proportion of keypad completes with no errors is around 64 percent, which compares to only 13 percent for speech. At the other extreme, the proportion of cases with five or more errors is only around 1.5 percent for keypad compared to almost 19 percent for speech.

Table 5-4. Proportion of cases experiencing errors (time out / invalid response) in NCVS 1 by response mode and completion status

	None	1 or more	5 or more	10 or more
Keypad[1]	62.1%	37.9%	1.7%	0.1%
Completes Only	63.9%	36.1%	1.5%	–
Partial Completes Only	46.3%	53.7%	2.7%	0.7
Speech[2]	12.15%	87.9%	21.6%	2.2%
Completes Only	13.1%	86.9%	18.7%	1.8%
Partial Completes Only	6.6%	93.4%	39.1%	4.6%

[1] Total cases for keypad response mode includes 1,450 cases.
[2] Total cases for speech response mode includes 1,350 cases.

Table 5-5 shows the percentage of respondents encountering specific types of errors, including timeouts, invalid entries and requesting help for NCVS 1. For keypad mode respondents, timeout errors were the most common source of error. This likely reflects instances where the respondent needed more time to think about their response, or didn't hear the question well enough to provide a response. Invalid entries for the keypad were rare with only 1.2 percent of all respondents encountering at least one error of this type.

Table 5-5. Proportion of cases experiencing errors and request for help in NCVS 1 by error type and response mode

	None	1 or more	5 or more	10 or more
Keypad				
Timeout	67.2%	32.8%	0.8%	0.1%
Invalid entry	98.2%	1.2%	0.1%	–
Help request	89.0%	11.0%	–	–
Speech				
Timeout	47.0%	53.0%	3.6%	0.2%
Invalid / unrecognized	23.1%	76.9%	9.5%	0.4%
Help request	74.2%	25.9%	0.4%	–

[1] The total number of cases for keypad respondents is 1,450, for respondents assigned to speech entry the total is 1,350..

In contrast, the speech mode has high rates of error for both timeouts and invalid entries. Timeouts occurring for speech mode respondents are likely to have similar causes as keypad respondents. However, the larger proportion of cases with at least one error compared to keypad responses (53.1 vs. 32.8 percent) likely suggests additional causes. Unlike the keypad entry mode, the speech system is dependent on detecting a response. If the utterance (response) provided by the respondent is too low, the system will continue to wait until the allotted time for a response has passed. By the time the respondent has realized their response was not picked up, the system will have already started the error prompt to replay the question.

Invalid entries for speech mode are extremely high. There are several reasons why these might have occurred. The speech system must be able to detect any utterance (sound/response) and therefore is susceptible to any sound that is detected. If there is too much ambient noise, additional utterances outside the expected response are provided, or if there are any other disfluencies the system will pick this up and have difficulty matching to an appropriate response. The substantial decrease from about 77 percent for one or more errors to 9.5 percent for five or more errors indicates that respondents were able to learn how to provide responses that were understandable by the IVR speech system. Nonetheless, the fact that this type of error was so common could be one reason why completion of the survey was lower for this mode of entry.

Speech respondents were more likely to request help than keypad respondents (25.9 percent vs. 11.0 percent). This is another indication that the respondents had more difficulties with the speech application.

Table 5-6 provides the proportion of cases by frequency of errors and response mode for NCVS 2 only. As was observed for NCVS 1, the proportion of errors was much higher for the speech response mode than for the keypad response mode. Another striking observation is the increase in one or more errors for the keypad mode compared to NCVS 1 (37.9 percent for NCVS 1 vs 62.0 percent for NCVS 2). For speech mode respondents the proportion with one or more errors was relatively the same, but the proportion with five or more errors increase dramatically almost doubling (21.6 percent: NCVS 1 versus 40.9 percent: NCVS 2). Given the added demands of NCVS 2, in terms of greater detail, and added length, it is possible that respondent fatigue is contributing to the increase in error incidence.

Table 5-6. Proportion of cases experiencing errors (time out / invalid response) in NCVS 2 by response mode

	None	1 or more	5 or more	10 or more
All Cases[1]	26.7%	73.3%	20.6%	4.5%
Keypad	38.0%	62.0%	4.3%	0.5%
Speech	12.7%	87.3%	40.9%	9.4%

[1] Includes any case that completed at least one incident in NCVS 2 and represents a count of respondents rather than incidents. The total number of respondents for which error data was available was 1,006. For keypad response mode the total number of respondents was 558 and for speech response mode 448.

5.3 Respondent Perceptions

Measures of respondents' perceptions of the IVR interview were measured with a series of debriefing questions that were asked at the conclusion of the interview. The respondent was not asked these items if there was a break off. If the respondent did not report any victimizations, they were skipped to this section. If a victimization was reported, then all of the NCVS-2 forms would have to be completed before this section was administered.

Tables 5-7 through 5-12 provide the results for each of the six measures collecting respondent perceptions of the IVR interview. Overall, most respondents reported being satisfied with the questionnaire. Across both sample frames, respondents assigned to the keypad mode generally provided higher ratings of satisfaction than those assigned to the speech mode. Those who did not report a victimization provided higher ratings of satisfaction.

Table 5-7. Perceived satisfaction with IVR questionnaire

	ABS Frame		Police Frame	
Satisfied with questionnaire?[1]	Keypad	Speech	Keypad	Speech
All Respondents	81.8 (979)	68.5 (879)	74.1 (290)	64.5 (231)
No Victimizations	84.0 (724)	69.3 (690)	83.7 (104)	71.3 (94)
w/Victimizations	75.7 (255)	65.6 (189)	68.8 (186)	59.9 (137)

[1] Sum of top two categories (very satisfied and satisfied) for these items from a five point satisfaction scale.

Table 5-8 provides the proportion of respondents who said they had difficulty understanding the IVR system. This does not seem to be a problem for the IVR, with a relatively small percentage of respondents reporting this type of issue.

Table 5-8. Perceived difficulty understanding the IVR system

	ABS Frame		Police Frame	
Problems understanding IVR?	Keypad	Speech	Keypad	Speech
All Respondents	5.6 (949)	2.7 (846)	6.3 (287)	2.6 (229)
No Victimizations	5.0 (694)	2.7 (660)	6.9 (102)	3.3 (92)
w/Victimizations	7.1 (255)	2.7 (186)	5.9 (185)	2.2 (137)

Table 5-9 provides the proportion of respondents who experienced difficulty with the IVR system understanding them. This only includes respondents assigned to the speech response mode. The proportion of respondents experiencing difficulty being understood by the IVR system is much higher for those reporting a victimization ($p<.05$). This might be attributable to interview length, as the longer the interview the more opportunities the respondent has of encountering difficulty.

Table 5-9. Perceived difficulty of IVR system understanding the respondent (Speech only)

Did IVR have trouble understanding R?	ABS	Police
All Respondents	34.4 (848)	32.9 228
No Victimizations	32.1 (661)	26.1 (92)
w/Victimizations	42.8 (187)	37.5 (136)

Table 5-10 provides the proportion of respondents who rated using the keypad to enter responses as very easy or easy. This only includes respondents assigned to the keypad response mode. Overall, using the keypad was viewed as easy by a vast majority of respondents.

Table 5-10. Perceived difficulty responding using telephone keypad (keypad only)

How easy was to respond using keypad?[1]	ABS	Police
All Respondents	96.6 (948)	94.8 (271)
No Victimizations	97.4 (693)	93.1 (102)
w/Victimizations	94.5 (255)	95.7 (184)

[1] Sum of top two categories (very easy and easy) for these items from a five point satisfaction scale

Table 5-11 provides the proportion of respondents who found any part of the NCVS IVR as confusing or frustrating. Between keypad and speech respondents the proportions are relatively the same for respondents reporting no victimizations. Low levels of perceived confusion or frustration with the IVR questionnaire are observed. However for respondents reporting (and completing) a victimization the rates increase. For the ABS frame, the rates more than double to 23 percent for keypad respondents and 25 percent for speech respondents. For the police frame a similar increase to 25 percent is observed for speech respondents, but for keypad respondents the increase to 21 percent is not as great. This is due to rate of nearly 19 percent for respondents not reporting a victimization.

Table 5-11. Perceived difficulty of confusion with IVR questionnaire

	ABS Frame		Police Frame	
Anything about questionnaire confusing or frustrating?[1]	Keypad	Speech	Keypad	Speech
All Respondents	14.0 (947)	12.0 (839)	20.3 (286)	18.9 (228)
No Victimizations	10.7 (693)	8.3 (652)	18.6 (102)	8.7 (92)
w/Victimizations	23.2 (254)	25.1 (187)	21.2 (184)	25.7 (136)

[1] Sum of top two categories (very easy and easy) for these items from a five point satisfaction scale

Table 5-12 illustrated whether respondents felt they had too much, too little, or just the right amount of time to provide a response. Respondents were given 6 - 10 seconds to provide a response, depending on the question. Overall, a large majority of respondents felt the amount of time was just right.

Table 5-12. Perceived amount of time to respond

	ABS Frame		Police Frame	
Amount of time for response?[1]	Keypad	Speech	Keypad	Speech
All Respondents (total n)	(945)	(823)	(286)	(216)
Too Much	12.3	8.6	10.5	9.3
Too Little	5.0	1.5	7.0	1.9
Just Right	82.8	89.9	82.5	88.9
No Victimizations	(692)	(640)	(102)	(87)
Too Much	11.3	8.4	3.9	4.6
Too Little	4.5	0.9	6.9	1.1
Just Right	84.2	90.6	89.2	94.3
w/Victimizations	(253)	(183)	(184)	(129)
Too Much	15.0	9.3	14.1	12.4
Too Little	6.3	3.3	7.1	2.3
Just Right	78.7	87.4	78.8	85.3

[1] Sum of top two categories (very easy and easy) for these items from a five point satisfaction scale

5.4 Other Issues

There were three other issues that are related to the usability of the IVR. One was discussed in Chapter 3 related to the income question. The format adopted for this study asked respondents to directly key in their income. While this potentially yields precise data, it resulted in two problems. One was that a number of respondents refused to answer the question by keying in '0'. A second problem was that a number of respondents did not key in their data correctly. This was evident by the high proportion of individuals who had extremely low incomes. We suspect respondents either

lost track of the number of digits they had entered or just entered their data in some other unit (e.g., 1,000s). The income question could be improved by using a series of unfolding categories, much like that used on the CATI. This would not require any entry of numbers and it would be less sensitive to ask. Both of these improvements would lead to a decrease in the amount of missing data.

A second issue was the administration of a question where the respondent should select a single response from a list. Turning these into 'yes/no' responses lead to some missing data when respondents answered 'no' to all of the options. For example, for the tenure question, respondents were asked about each residence type separately and could say no to all options. For the IVR Telephone group, this resulted in a large proportion of respondents having an 'unknown' housing type. This can be avoided by either presenting short lists, with numbers preceding each option and asking the respondent to enter the number. In our case, this would have led to a large number of speech interviews using the keypad (to enter the number). Another solution, one which we used on selected questions, is to have a check to make sure the respondent selects a category. If they do not, then they can be presented with an 'other' category or they are re-asked the question.

A third issue concerns the accuracy of the coding from the IVR system. There is some error when answers are recorded. This error is more prevalent for the speech application. Speech recognition is not a perfected technology, at least not when used in a survey environment where there can be a lot of ambient noise and uncertain telephone quality. For example, Johnston et al (2013) report an error rate of between 4 percent and 11 percent, depending on whether the system verified responses. The vocabulary used on this project's IVR was highly restricted, compared to the Johnston example. The use of Yes/No responses will result in fewer recognition errors. Nonetheless, our review of audio recordings did uncover errors. On the screening instrument, those heard by analysts primarily resulted in a 'no' being misinterpreted as a 'yes' response. If this occurred at a screener item, it would attribute a victimization to someone who did not intend to report it. This type of error would eventually be caught when the respondent was asked to verify the occurrence of an incident at the beginning of the detailed incident form. Error on the NCVS-2 would not be caught by the respondent. If this error occurred where there is a skip pattern, the respondent may go down the wrong path. This will lead to measurement error (e.g., misclassification of the incident).

While the keypad mode is more accurate than speech, it is also subject to a small amount of error. This occurred when respondents either miskey an answer (see discussion above on income) or the keys are hit more than once. In the latter case, the IVR advances two questions, or in the case of a 'mark all that apply' question, will be coded as having an extra response category. This occurred, for

example, on the race question where some individuals were coded as 'multi-racial' because an extra race category was entered in error.

This project did not complete a systematic review of the questionnaire. However, we did review all crime summaries at the end of the incident form and compare these summaries to the TOC assignment (see Chapter 7). We believe this review caught many of the entry errors that resulted in an inaccurate recording of a crime. But there are still errors that remain in the data-set.

5.5 Summary of Usability of IVR

The average time to complete the NCVS-1 screener ranged from seven minutes to 10 minutes, depending on whether a victimization was reported. As evidenced by the response rate results, respondents completed this portion of the interview without many dropping out. Once a victimization is reported, however, the time to complete goes up to between 16 and 24 minutes. As pointed out in earlier chapters, respondents start dropping out when getting to NCVS-2. This is reflected in the respondent perceptions of the interview. Those who had at least one victimization were more likely to report being less satisfied with the questionnaire, confused, frustrated and that the survey was too long. Note that these are the respondents that completed the entire survey, so these perceptions likely underestimate the problems all respondents had who tried to complete a detailed incident form.

There are two ways one might reduce the higher dropout rate for the NCVS-2. One would be to include the keypad entry mode (see further discussion below) and an incentive. Both will increase the number of persons that complete the survey. Second, the NCVS-2 should be adapted more specifically for IVR administration. This involves simplifying and shortening the instrument. One way to do this would be to reduce the amount of detail in the TOC assignments, for example ask questions to classify events into the major types of crimes, rather than the current two-digit coding scheme. It might also be possible to simplify the respondent's task by using more open-ended questions to collect the data. Respondents could speak their responses into the IVR, as they did successfully when providing summaries. For example, an open ended response might be used for the location of the incident or the type of item stolen. Use of verbatim responses will add costs when coding the data and it may not consistently provide all of the details necessary to classify an incident. The open-ended questions would also have to be placed so that skip patterns can be programmed.

The speech respondents had the highest error rates, with the driving problem being the system failing to recognize their responses. On NCVS-1, 76 percent of the speech respondents had at least one question where the system did not recognize their response. Interestingly the perceptions of the speech respondents expressed in the debriefing questions did not reflect these problems. We suspect that this may be because those respondents who became frustrated dropped out before the debriefing questions were administered, as evidenced by the lower response rates for these individuals.

There were also several problems with the design of selected IVR questions, including income and some of the 'mark all that apply' items. We make recommendations above on how to remedy these issues in a redesign of the questionnaire.

Finally, there were instances where the IVR system, primarily for the speech respondents, did not correctly code the data. In combination with the issues that speech respondents have with the system recognizing their responses, we recommend that future IVR applications rely either on a keypad response mode or allow respondents to choose which mode they would like to use. Primary reliance on speech should be considered as the speech recognition technology evolves. Reliance on the keypad mode in the short term should significantly reduce the level of frustration of respondents and increase overall data quality in terms of response rates and data entry.

Adapting the NCVS-1 6

Chapter Highlights and Key Findings

- Approximately 38 percent of the IVR respondents reported at least one victimization. This differed significantly by sample frame (30 percent for ABS, 71 percent for Police). It also differed by the method of contact (mail > telephone), even after controlling for household and demographic characteristics.
- 32 percent of those reporting a victimization also reported a duplicate incident.
- There was no difference in the proportion reporting a victimization between the IVR and CATI.
- The mean number of victimizations reported on the IVR was high. For the ABS sample, this differed by mode of contact (mail > telephone). Of those in the ABS sample that reported a victimization, 40 percent reported more than one victimization and ten percent reported four or more.
- The mean number of victimizations significantly differed between the IVR and the CATI.
- About 10 percent of those reporting a victimization modified their answer when asked to verify their reports at the beginning of NCVS-1. Many reduced the number they originally reported, with about half reporting that it did not occur at all.
- Take-away point #1: The IVR can effectively screen for victims. With respect to the proportion of persons that report a victimization, it is similar to the CATI.
- Take away point #2: The procedure to unduplicate incidents was successful. It caught a significant number of individuals who reported duplicate events and there wasn't any evidence of duplicates from the detailed summaries (next chapter).
- Take-away point #3 : The victimization rates from the IVR screener were high when compared to the CATI and to the ongoing NCVS. While the procedure used to verify the incidents at the beginning of the incident form did reduce this number somewhat, it may not have caught all. We recommend putting in more explicit verification of the date of the incidents when it is initially reported on the screener.

This chapter continues the discussion of whether the NCVS can be adapted for IVR administration by examining the responses to NCVS-1. The IVR adapted the NCVS-1 by breaking up the cues into different subsets. After a 'yes' response to a subset of cues, the program asked for a summary

of the incident(s). The program then skipped to the next screener item. In addition, the IVR used a specialized procedure to unduplicate events that re reported multiple times on the screener.

In this chapter analyses are presented that examine how the above adaptations worked for NCVS-1. The first section presents the overall prevalence rates on screener and how these rates varied by the different experimental conditions. The second sections reviews how often reports were unduplicated. The third section compares results to the CATI. The final sections examine the overall level of victimization and the verification procedure when transitioning between NCVS-1 and NCVS-2.

6.1 Response to the Screener

Table 6-1 provides the percent of respondents that reported a victimization at each of the screener question. A total of 38 percent of respondents reported a victimization to at least one of the screener items. As with the NCVS, the first questionnaire item on theft has the largest number of respondents reporting a crime. The percent reporting a crime goes down substantially after that item. There is a large difference between the ABS and Police Sample frames, with almost three-quarters of the respondents from the police frame reporting a crime compared to one-third of the ABS respondents.

Table 6-1. Percent of IVR Respondents Reporting a Crime at the Screener by Screener Item and Sample Type*

		Sample Frame		
Items	All (n=2553)	ABS (n=2017)	POL (n=482)	p-Value
Total	38.1	30.3	71.4	0.0001
Item 36a - theft	25.9	19.1	55.0	0.0001
Item 37a - Burglary	7.0	5.5	13.1	0.0001
Item 39a - Motor vehicle	6.2	4.4	13.9	0.0001
Item 40a - Stealing	6.7	4.9	14.5	0.0001
Item 41a - Attack	2.3	4.9	14.5	0.0001
Item 42a - Nonstranger Assaults	2.1	1.7	3.7	0.006
Item 43a – Unwanted sex	1.0	0.8	1.7	0.092
Item 44a-45a - Police	1.9	1.5	3.5	0.0043

* Excludes: 1) all addresses offered an incentive and 2) all addresses that did not match to a telephone number

The percent of respondents reporting a victimization also varied by the mode of contact (Table 6-2). At an item level, the most significant differences are for Burglary (item 37a) and Stealing (item 40a). Some of this difference may be due to differences in the socio-demographic composition of these

two groups. As noted in Chapter 3, there are significantly more older people in the IVR Telephone mode. We tested whether the differences by mode of contact differed once controlling for age. This was done with a logistic regression predicting the proportion victimized with mode of contact and demographic characteristics of the respondent. Mode of contact remained significant in this regression.

Table 6-2. Percent of IVR Respondents Reporting a Crime at the Screener by Screener Item and Mode of Contact*

	Mode of Contact		
Items	Mail (n=2095)	Telephone (n=458)	p- Value
Total	39.2	32.7	0.01
Item 36a - theft	26.5	23.1	0.13
Item 37a - Burglary	7.7	3.5	0.001
Item 39a - Motor vehicle	6.5	4.6	0.11
Item 40a - Stealing	7.2	4.4	0.03
Item 41a - Attack	2.5	1.5	0.21
Item 42a - Nonstranger Assaults	2.3	1.1	0.09
Item 43a – Unwanted sex	1.1	0.22	0.07
Item 44a-45a - Police	1.9	2.2	0.64

* Excludes: 1) all addresses offered an incentive and 2) all addresses that did not match to a telephone number

One possible interpretation of this difference is the tendency for individuals who have been victims of crime to be more interested in the survey and more likely to respond by mail. A mail contact is likely to be the most subject to this type of "avidity" bias, all other things being equal, because the respondent is reacting to written material sent to the household. While a respondent to the telephone survey may have read the advance material, they are also reacting to the interviewer's introduction about the survey, which may evoke other reasons for the individual to participate on the survey.

Tabulations were also run by the incentive condition and mode of entry, neither of which were significant at the 5 percent level.

6.2 Unduplication Procedure

The average number of positive screener items reported was .9. This differed significantly by sample frame, with the ABS respondents reporting an average of around .6 and the Police frame reporting an average of 1.7. When responding 'yes' to a screener item, the program checked if the person had reported 'yes' on a prior screener question. If so, the respondent was asked if the incident was the

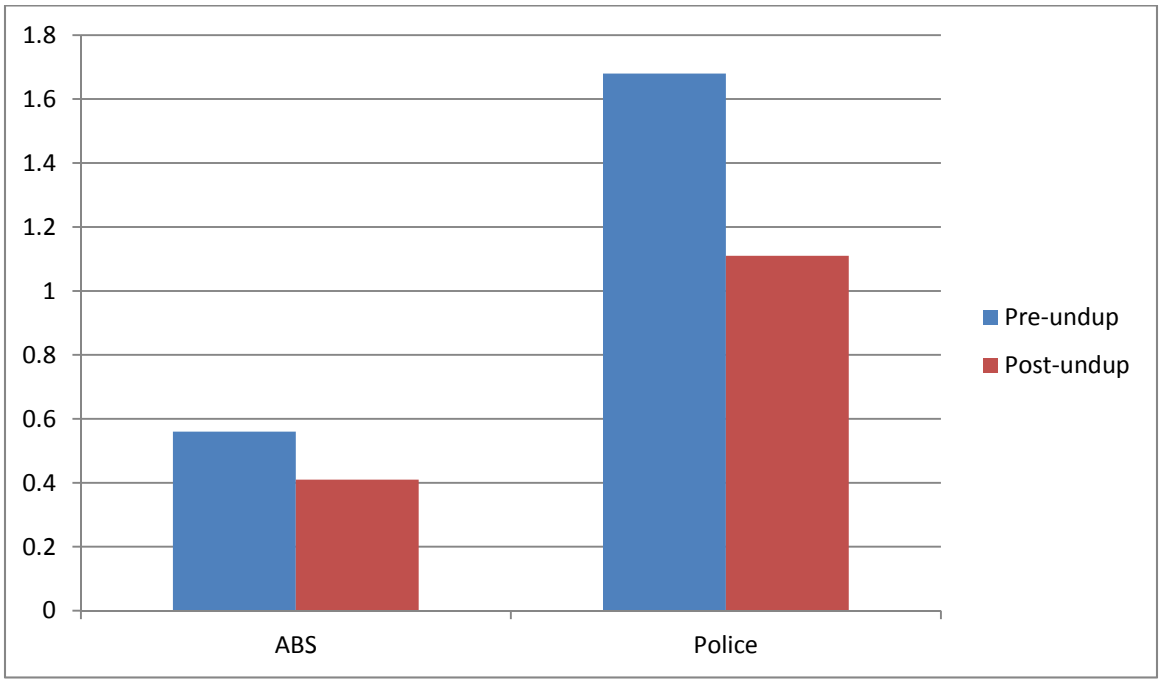
same as something that had already been reported. If it was, the respondent was skipped to the next cue and no details were collected on the victimization.

Overall, about 32 percent of the respondents that reported a crime also reported at least one duplicate incident. Duplicate incidents may have occurred because the respondent did not attend or listen to the exclusionary phrase at the beginning of each screener item. For example, if a victimization was reported on item 36a (theft), all of the remaining questions would begin with the phrase "Other than any incidents already mentioned….". Respondents may not have always heard this phrase. A second reason for duplicates may be the way the screener items were structured. If more than one set of cues were read for a particular item (e.g., item 37a), the exclusion language was not included in the second or subsequent cues. This may also have led to reporting duplicates. When reviewing which subsets of cues were caught as duplicates, both of the above explanations are apparent. Some duplicates were found when the phrase was read, while other duplicates occurred when it was not included.

Given the high rate of unduplication, the procedure seemed to work reasonably well. Review of the summary reports that were provided on NCVS-2 did not reveal a significant number of incidents from separate screener items that were part of the same event.

Once unduplicating the screener items, the number of positive responses goes down significantly (Figure 6-1).

Figure 6-1. Mean number of screener items before and after unduplication by Sample Frame

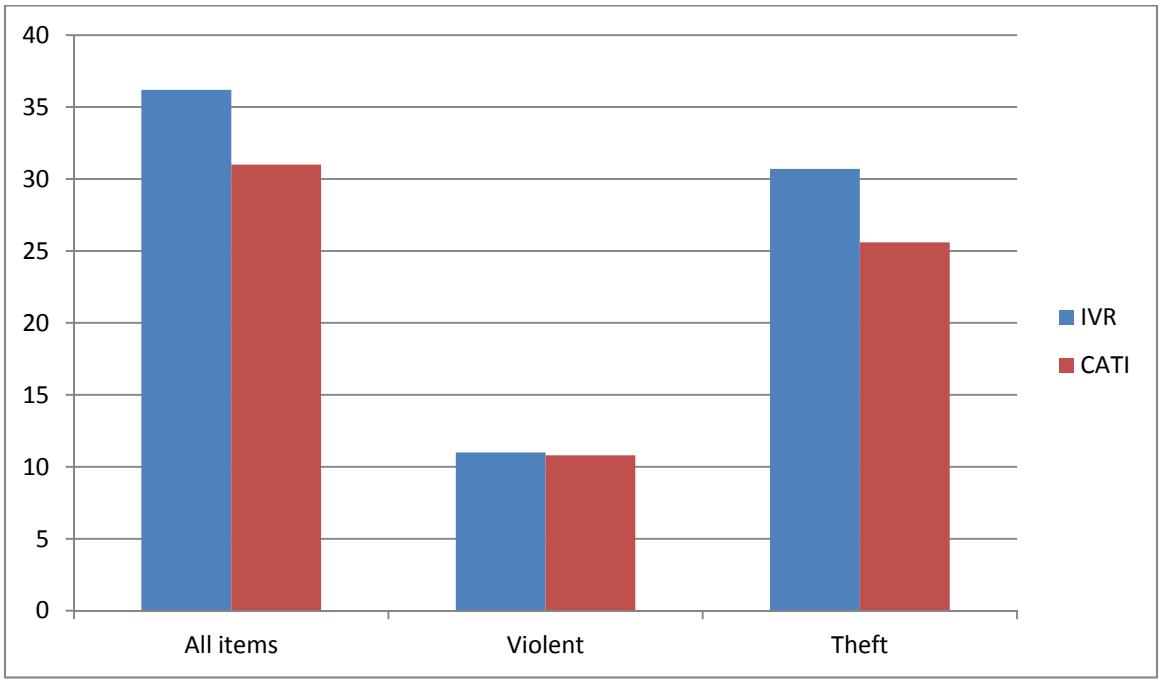


6.3 Comparison of IVR to CATI

Another way to benchmark the performance of the IVR screener is to compare it to the CATI. Since the CATI has an interviewer administer the items and collect the relevant details, it provides a benchmark that is closer to the ongoing NCVS. The CATI screener was administered using procedures that were closer to the NCVS. When reading each screener question, all of the cues within the item were read before asking for a summary of the incident(s). This is in contrast to the IVR which collected summary information immediately after a 'yes' response and then skipped to the next screener item. The CATI did differ from the NCVS in one way ---- the screener cues were broken up into sets in the same way as on the IVR version. For each set, respondents were asked to answer 'yes/no'. After all sets were read for the screener item, the interviewer asked for a summary of any incidents that were reported.

The IVR does have a slightly higher proportion of individuals reporting a victimizations at the screener items on theft when compared to the CATI (Figure 6-2). However this difference is not statistically significant. This result holds even after controlling for household and demographic characteristics in a logistic regression..

Figure 6-2. Proportion reporting a crime on NCVS-1 by Mode of Interview*



* No statistically significant differences between modes. IVR excludes respondents assigned to the incentive experimental condition and to addresses that did not match to a phone number.

6.4 Number of Victimizations Reported on NCVS-1

For each positive response to a screener item, the respondent is asked how many times the incident occurred. This number is used to guide the number of summary reports that are requested and to the number of detailed incident forms that are requested. The mean number of victimizations for the IVR is quite high (Table 6-3). Overall, the mean is .9, with a much larger mean for the Police sample (2.04) than the ABS (.63). But even for the ABS, this mean would imply a rate of approximately 630 per 1,000 population, much higher than on the NCVS (which was around 160 in 2011). The mean victimization rates differ significantly by mode of contact for the ABS sample. The IVR Mail is significantly higher than the IVR Telephone, consistent with the differences observed for the prevalence rates reported earlier.

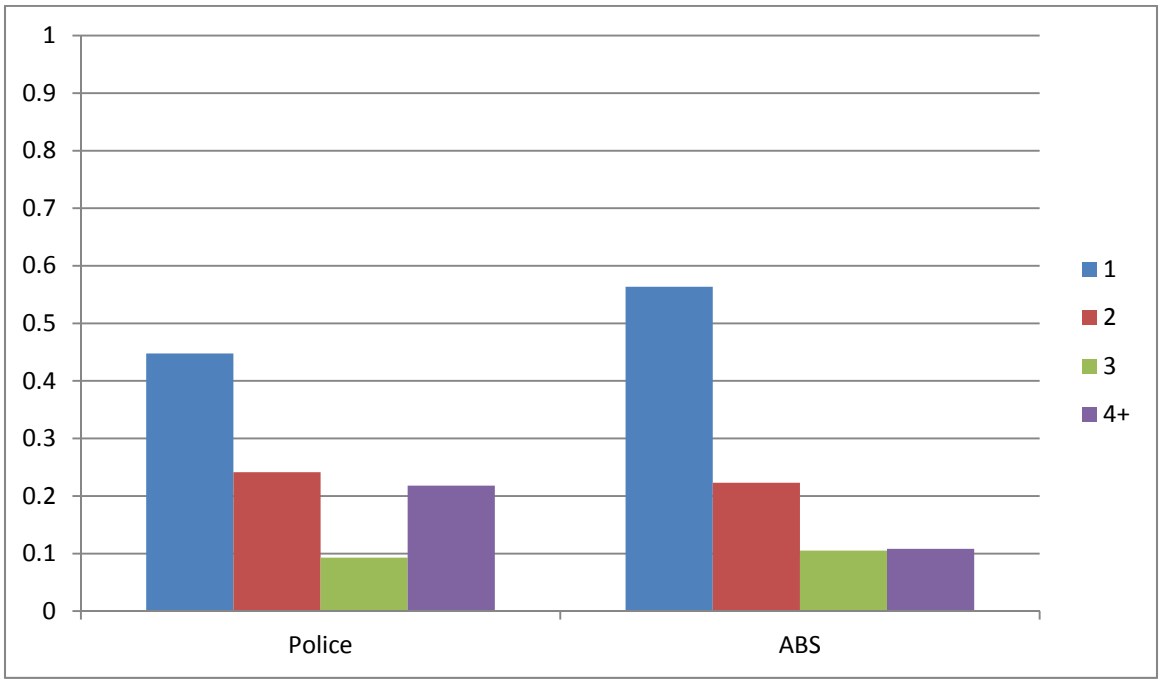
Table 6-3. Mean number of victimizations reported on the IVR screener by Mode of Contact and Sample Frame*

	N	Mean	Standard Error	Range
Total	2553	0.90	0.05	0-53
Mail Contact	2095	0.93	0.05	0-53
Telephone Contact	458	0.75	0.12	0-37
ABS+	2071	0.63	0.04	0-9
Mail Contact	1699	0.67	0.05	0-53
Telephone Contact	372	0.47	0.09	0-23
Police	482	2.04	0.15	0-37
Mail Contact	396	2.10	0.15	0-26
Telephone Contact	86	1.97	0.53	0-37

* IVR excludes respondents assigned to the incentive experimental condition and to addresses that did not match to a phone number
+ Difference between mail and telephone is significant at p<.05

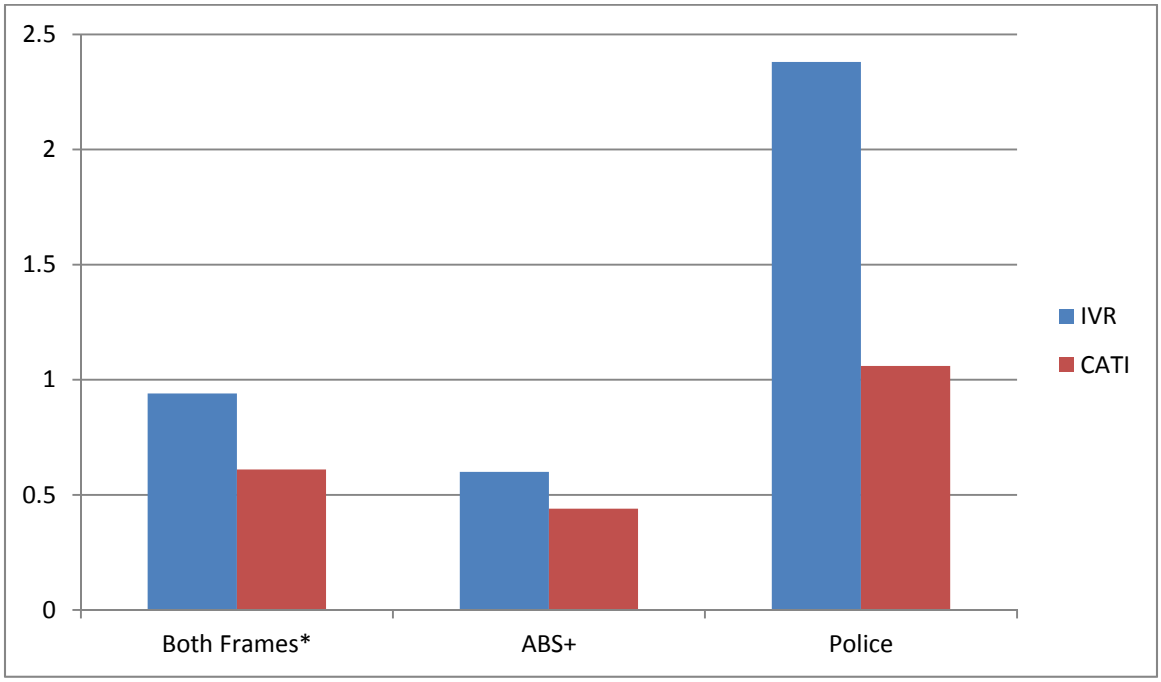
Figure 6-3 provides the distribution of the number of victimizations. This collapses reports of four or more incidents into a single category. As indicated by the above discussion, there are quite a few individuals who reported more than one incident. For the police sample, 20 percent of the sample reported four or more incidents. This is not as dramatic for the ABS sample. But even for this group a little less than half reported two or more incidents.

Figure 6-3. Proportion Reporting 1 or More Victimizations by Number Reported and Frame



There were significantly more reports of victimizations on the IVR when compared to the CATI (Figure 6-4). The difference for ABS is statistically significant. While the comparison for the Police Frame is very large, the number of sample cases in the CATI mode is too small to yield a statistically significant result.

Figure 6-4. Mean number of Victimizations Reported on NCVS-1 by Mode of interview and Sample Frame+



* Difference is significant at p<.05; + Difference is significant at p<.10
+ IVR excludes: 1) addresses where an incentive was offered and 2) addresses that did not match to a telephone number

6.4 Verifying Incidents and Transitioning to NCVS-2

One of the reasons why the number of victimizations may be high for the IVR is the tendency to report out-of-scope incidents. An interviewer is in a position to verify basic eligibility criteria before the incident is recorded on the screener. For example, an interviewer can align the date of occurrence and the eligibility of the victim with NCVS criteria. As noted in chapter 2, the IVR had a procedure that asked respondents to verify that the incident was within the reference period:

> "You said that during the last 12 months {TYPE OF INCIDENT} occurred {NUMBER OF TIMES}. Is that correct?"

If the respondent answered 'no', they were asked how many times the incident occurred. The respondent could then key in a '0' if the incident did not happen or key in a different number if appropriate. If a '0' was entered, the IVR went to the next screener item for which a 'yes' response had been given or to the end of the interview.

Overall, about 90 percent of the respondents confirmed their responses (see shaded diagonal of Table 6-4) and 10 percent used this opportunity to change their answer. Most of the changes were to lower the number of victimizations (see above diagonal in Table 6-4). This was to presumably take out ineligible events (e.g., outside the reference period). About half of those that changed, nullified all of their events by reporting '0' (5.8 percent out of 10 percent).

Table 6-4. Initial number of reports of Victimizations on IVR for NCVS-1 by Number Reported after verification (Percent of those initially reporting a victimization)

	Initial Report				
Corrected Number	1	2	3	>=4	Total
0	3.6	0.8	0.7	0.6	5.8
1	47.9	0.7	0.4	0.1	49.2
2	0.3	21.2	0.6	0.9	23.0
3	0.4	0.1	8.2	0.9	9.7
>=4	0	0.1	0.1	12.1	12.3
Total	52.3	22.9	10.1	14.7	100
N	508	0	98	143	972

+ Excludes: 1) addresses where an incentive was offered and 2) addresses that did not match to a telephone number

6.5 Summary of Adapting NCVS-1 for IVR

Adapting the IVR screener was reasonably successful. About one-third of the respondents reported at least one duplicate victimization, despite the exclusionary phrase inserted at the beginning of each screener item. This indicates some success for the verification procedure used on the IVR. Some of the duplication may have been due to the way the screener questions were broken up, with later sub-items not including the exclusionary phrase. But it is also likely that at least some respondents were not listening carefully and reported the same events, even when the exclusionary phrase was included. Review of the summaries provided for the incidents confirms that most, if not all, of the duplicate incidents were eliminated by this procedure. Consequently we believe the procedure was successful.

There was a small, but significant difference between the two modes of contact, with the IVR Mail having a higher proportion reporting victimizations than the IVR Telephone. This result holds up after controlling for differences in household and demographic characteristics. When comparing to the CATI, the IVR had a nominally higher proportion of persons reporting a victimization but the difference was not statistically significant. This last result confirms that at least for screening for victims, the IVR seems to function about like the CATI.

The number of victimizations reported for the IVR was high. For the ABS sample, a little less than half of those that reported a victimization reported at least two. While there was a tendency for respondents to reduce these numbers when verifying at the beginning of NCVS-2, this only affected about 10 percent of the respondents reporting a victimization. The number reported on the IVR was significantly higher than the CATI and was much higher than what is reported on the ongoing NCVS.

This raises the possibility that the IVR reports may contain additional ineligible events. One change that should be implemented is to have a more explicit confirmation of the month/year of each incident reported. The confirmation used at the beginning of the detailed incident form may have been too global. It combined verification of both the date and the number of times. It may be that respondents were not concentrating on the dates of each of the events they initially reported. They also may not have been thinking carefully about the reference period. It would be preferable to collect the month and year for each incident at the time the incident is initially reported on the screener. This will force the respondent to think about the timing of the events relative to the reference period.

Adapting the NCVS-2 for the IVR 7

Chapter Highlights and Key Findings

- For the ABS Sample, about 10 percent of those reporting a victimization were capped at three incidents. For the Police Sample about 20 percent were capped.
- About 20 percent of the incidents could not be classified. Of those that could be classified, about 15 percent were classified as violent crimes and 85 percent were classified as property crimes.
- About 90 percent of the incidents classified into a TOC were confirmed as eligible by examination of the summary provided by the respondent. About one third of the eligible incidents were not classified correctly. Of those incidents that did not have a TOC assigned, about 40 percent were actually an eligible crime.
- One type of error that lead to incorrect TOC assignment was respondent confusion about: 1) who is an eligible victim (e.g., non-household members), 2) the reference period and 3) which incidents are being reported on NCVS-2.
- A second type of error was related to filling out key items on NCVS-2. The most prevalent error was the failure to report a theft when asked (Q88 and Q89 on NCVS-2). Other errors related to incorrectly answering key NCVS-2 items and not getting routed on the right path.
- Take-away point #1: Adapting the NCVS-2 for the IVR is possible. The success in classifying many crimes, at least in the major crime categories, is promising. For example, the IVR compared favorably to the CATI application. However several of the above issues need to be addressed through adapting the NCVS to an IVR application.
- Take away point #2: One suggested change to an IVR version is to collect more information on the NCVS-1 about the month and year of the event. This should reduce the amount of external telescoping and provide a way to reference events when linking to NCVS-2.
- Take-away point #3 : A second suggested change is to adapt NCVS-2 to a keypad version of the IVR. One such change would use numbered lists, rather than restricting questions to a yes/no format. A second change would be to record verbatim responses for lengthy items (ie what was stolen), rather than asking in closed-ended questions.
- Take-away point #4: If it is desirable to move forward with the IVR for NCVS-2, more detailed research with the IVR data-set should be completed which embellishes the quality analysis described in this chapter. This involves transcribing the summaries provided on NCVS-1. This will add more information on possible errors, as well as suggest ways to modify the questionnaire for the IVR.

This chapter evaluates the performance of the IVR version of the NCVS-2. The IVR implemented an abbreviated version of the NCVS-2 by only including those items necessary to classify events into a two-digit TOC. On the NCVS, the interviewer performs two critical functions that need to be replicated by the IVR. One is to sort through the summary reports from the screener, decide which ones are eligible for a detailed incident form and orient the respondent to each incident when asking the NCVS-2 questions. The IVR mimicked this procedure by referring to the screener item that generated the report when starting the incident form. A second critical function is ask open-ended questions on NCVS-2 and code the answer into a pre-designated category. There are also a number of verification items that intervewers do not ask if they already know the answer. The IVR adapted these questions to a 'yes/no' format, requiring respondents to answer all questions necessary to complete survey.

In this chapter we discuss how well the IVR performed these tasks, by assessing how well the NCVS-2 classified incidents into one of the major crime categories. Before discussing this, the method used to cap the number of NCVS-2 incidents is described.

7.1 Capping the Number of Incidents

To minimize burden, the number of requests to fill out an NCVS-2 was limited to three. The capping used a logic which reviewed all of the screener incidents and the number of different screener items involved. When four or more incidents were reported and:

1. Four or more screener items were reported, the priority was to select one incident from each of the crime types with the highest priorities (Appendix D). If more than one incident was reported for a particular crime type, the most recent incident was selected.

2. Three screener items were reported, the program selected the most recent incident reported for each crime type.

3. Two screener items were reported, the program selected two incidents from the crime type with the highest priority (Appendix D).

4. One screener item was reported, take the three most recent incidents.

The count of incidents prior to verification at the beginning of NCVS-2 was used to decide on whether a cap was needed. For example, if a respondent initially reported four incidents and the verification reduced this to three incidents, the program would only ask the respondent to report on two other incidents.

Approximately 10 percent of the ABS respondents and 20 percent of the Police sample were capped in this way.

7.2 Review of the Detailed Incident Forms

Approximately 80 percent of the detailed incident forms had a report that could be classified into a two digit TOC category (Table 7-1). There were small differences between the ABS and Police sample. Relative to the ongoing NCVS, there are several patterns that stand out for the ABS sample. One is the relatively high percentage of incidents classified as Rape and Sexual Assault. It is actually higher than robbery and about twice the percentage on the NCVS. A second difference is the high number of burglaries compared to thefts. On the NCVS, the ratio between thefts to burglaries is considerably higher than shown below.

Table 7-1. Percent Distribution of Major Types of Crimes for IVR

Type of Crime	Total	ABS	Police
Rape & Sexual Assault	2.2	2.3	1.5
Robbery	2.5	1.0	4.8
Assault	9.3	8.6	10.4
Personal Theft	1.0	.9	1.3
Burglary	23.4	21.2	27.7
Theft	36.5	39.3	31.9
Motor Vehicle Theft	6.5	6.4	6.7
Not classified	18.4	20.0	15.8
Total	100.0	100.0	100.0
N	1262	782	480

In order to assess the quality of data provided on NCVS-2, all of the summaries provided by respondents at the end of NCVS-2 were transcribed and reviewed. On this review, the coder classified incidents into one of six categories (Table 7-2):

- Incident is out of scope (e.g., vandalism; nothing taken)
- Incident occurred to an ineligible person (e.g., neighbor)
- Duplicate incident with something already reported
- Eligible victimization but incorrect TOC
- Eligible Victimization
- Cannot discern from available information

Table 7-2. Percent Distribution of Quality Codes Assigned to Incidents by whether the incident was assigned to a 2-digit TOC code

Quality Codes Based on Summary	Total	Eligible TOC	Not Eligible TOC
Incident is out of scope	10.5	5.0	35.0
Ineligible person	2.5	1.6	6.5
Duplicate incident	1.6	1.8	.9
Eligible victimization, incorrect TOC	21.8	20.4	28.0
Eligible Victimization	40.3	49.3	0
Cannot determine	23.4	21.9	29.7
Total	100.0	100.0	100.0
N	1262	1030	232

This review was done by one senior person, with no systematic checks on the reliability of the coding. We also note that this coding relies on the summary given at the end of NCVS-2. There is additional information from the summary provided on NCVS-1 that could supplement this analysis if desired.

A significant number of incidents are classified as out of scope. Many of these were instances of vandalism or identity theft. Other examples are verbal altercations that do not rise to the level of an assault. Examples of summaries include:

> *"My debit card was used online to order items online. Called the bank and they returned my money"*
>
> *"A window was broken out of the car"*
>
> *"Threatening notes were put in my mailbox"*
>
> *"drug deal went bad and drug dealer chased someone firing three shots; one bullet went through the front door shattering glass"*

Also included in this category are incidents that were found to be out of the reference period.

The category for ineligible persons are incidents that either involved violent crimes that did not occur against the respondent or property crime that occurred against a non-household member. Duplicate incidents reflect confusion on the part of the respondent when sorting through multiple incidents. When repeating multiple instances of the NCVS-2, some respondents became confused

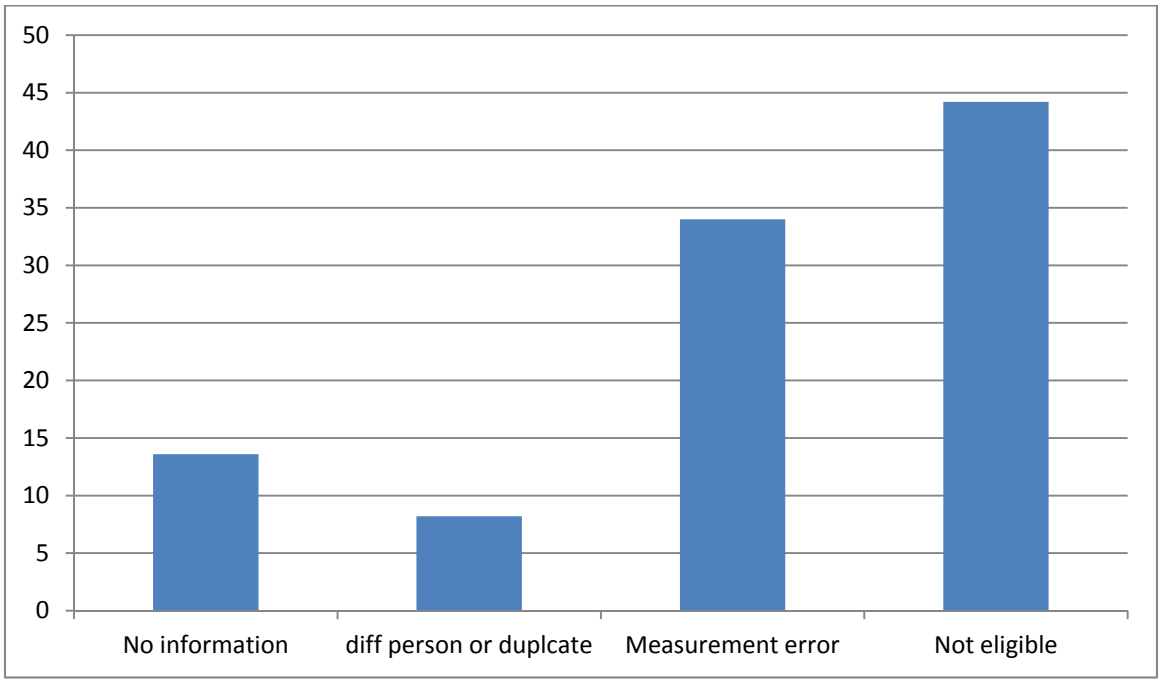
about which incidents they were supposed to report on. Some also combined reports about multiple incidents in the same reports.

In the case of the ineligible victims and duplicate reports, an interviewer would be able to prevent similar confusions. Additional checks on the screener may assist in reducing some of this problem. For example, it might be possible to confirm whether the victim resides in the household. Other checks could be incorporated into NCVS-2, once it is determined whether it is a violent crime.

A significant percentage of the incidents were misclassified, either by not receiving a 2-digit code or being given the wrong code. An incident was classified as in the wrong category if there was something in the summary that directly contradicted the classification. About one-third of the incidents could not be reviewed at all, either because no summary was given or the summary was just too ambiguous to determine if it was consistent with the coding. For both of these reasons, the estimates in Table 7-2 are likely to be underestimates of the error involved in the classification.

A second review was completed for all of the incidents that were not classified (Figure 7-1). The purpose of this review was to assess what caused the incident to not get a 2-digit code. In conjunction with the summaries used above, this provided some insight into whether the inability to classify was because the event was actually ineligible or there was a problem with how the NCVS-2 questions were answered. For example, if the incident indicated a burglary, we examined where the respondent veered off the path to classify the event as such.

Figure 7-1. Reasons why Incidents Were Not Classified into an eligible TOC code (n=232)



Almost 35 percent of these incidents are not classified because of an error. The vast majority of these (about 80 percent) are property-related where the respondent did not report a theft at the main gate questions to classify the event as a completed theft or attempted theft. The wording of these questions on the IVR was:

> "During the incident was something stolen or taken without permission that belonged to you or others in the household? If you are unsure, or don't know say 'don't know'. (adaptation of Q88 on NCVS)
>
> "Did the offender attempt to take something that belonged to you or others in the household? If you are unsure, or don't know say 'don't know'." (adaptation of Q89 on NCVS)

For the speech cases, we also reviewed the audio of the actual question/answer sequence to see if there were any signs of confusion or problems. Some of these cases involved non-native speakers who may have been confused by the questioning. Several other cases were individuals who selected the 'don't know' option to the question. Responding 'Don't Know' was associated with situations where there was nothing actually stolen, but the respondent was not willing to report an attempt. The plurality of these cases did not have a discernible reason for the respondent to answer 'no or

don't know' to both of the above questions. Examples of summaries for a few of these situations include:

"Person entered through the window with air conditioning ; stole tv, jewelry, deceased father's shoes, dvds, table boxes, tools, food, wine bottles, father's car, and other things; nothing was recovered"

"Happened at daughter's house; our vehicle was taken from her house; called police; sent detective to help; got vehicle back; daughter knew person who took it"

We have two conjectures on what may have occurred. One is the respondent was confused by the detailed questions and the initial summary they provided. Some respondents may have thought this question was asking about any additional items that may have been stolen, since they had already reported on the details. This type of error was found in the pilot of the Companion Study (Westat, 2013), which is interviewer administered. A second explanation is that respondents may have been distracted or not paying attention. Since these respondents would have already gone through a significant portion of the interview, they may have lost concentration or actually been responding 'no' to shorten the interview.

Many of the other types of problems were related to attempted burglaries. These fell into two general categories --- 1) the respondent was not routed correctly and not asked the pertinent question and 2) the respondent did not report an attempted break-in. Several examples of these include:

Wrong Routing

Tried breaking in through french doors. Items on evidence of forcible entry were not asked because R reported that the incident occurred near her home.

Back door was broken out - alarm went off and perp ran away. R reported that the offender did not try to get inside, so she was not asked variables about evidence of entry

Attempts

R stated that someone tried to get inside (by unscrewing the locks on the door) but reported that there was no evidence that the offender tried to get in by force

R stated that there was an attempted beak in on their house. They called the police and they dusted for fingerprints. However, R reported that the offender did not try to get inside

Using the summaries, a coder reviewed the incidents that were initially categorized as eligible victimizations but not classified correctly. The coder classified the event into one of the major type of crime categories. Table 7-3 provides a cross classification of the original TOC assignment and the revised TOC assignment. Note that some of the classifications do not change because the assignments were made at the major crime category level. Most of the changes occur within the two major classes of crime of violent and property. For Rape and Sexual Assault, 37 percent stayed in this category, with most being moved into the assault category. The summaries for these did not indicate any type of sexual assault. Only 18 percent of the burglaries originally classified stayed in this category. Most of these involve confusion about location, with the original incident occurring on the respondent's property, but no attempt to enter the home. Respondents reported these as occurring in the home.

Table 7-3. Percent of Original TOCs assigned in each Revised TOC for misclassified incidents

	Original TOC Classification							
Revised TOC	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Total**
1. Rape and sexual asslt	37%							2%
2. Assault	44%	50%			1%	1%		7%
3. Robbery		6%			1%			2%
4. Personal Theft	6%	6%			1%	8%		4%
5. Burglary			25%		18%	10%	10%	12%
6. Theft	6%	38%	25%		72%	65%	77%	62%
7. Motor Vehicle Theft	6%		50%	100%	6%	16%	12%	12%
Total	100%	100%	100%	100%	100%	100%	100%	100%
N	16	16	4	2	93	79	31	242

Many of the thefts were reclassified as either a burglary or a motor vehicle theft. In the case of burglaries, respondents had problems with the location variable or did not report any evidence of a break-in. The evidence question, for some respondents, was taken literally. For example, the question on evidence was

"Was there any evidence such as a broken lock or broken window?"

Several respondents answered 'no' to this even though they reported in the summary that their door was damaged (hinges taken out) or their door was kicked in.

For many that were re-classified into motor vehicle theft, the respondent answered 'no' to the question:

> "You said the offender vehicles or parts, such as a car, car parts, gasoline, other motor vehicles, or a bicycle or bicycle parts. Was it a car or other motor vehicle?"

This question is an example of problems with converting the open-ended question on the NCVS to a series of closed-ended yes/no questions. In this case, respondents may have gotten lost in the examples and did not hear the last part referring to a car or motor vehicle.

The opposite occurred for those originally classified as a motor vehicle theft. Most were re-classified as a theft. In almost of these cases, the car was broken into and an actual or attempted theft occurred.

7.3 Final Classification and Comparison to CATI

Table 7-4 provides a re-classification of the incidents using the re-coding described above. For those incidents that could not be reviewed because there was no summary information, the original classification was used. There is very little change in the percentage of crimes classified as a violent or property crime. Within Rape and Sexual Assault, there is a slight decrease in the percentage of incidents. The largest change is within the property crimes with a shift into the theft category. Many of these are coming from the previously unclassified incidents. Some are from those originally classified as a Burglary.

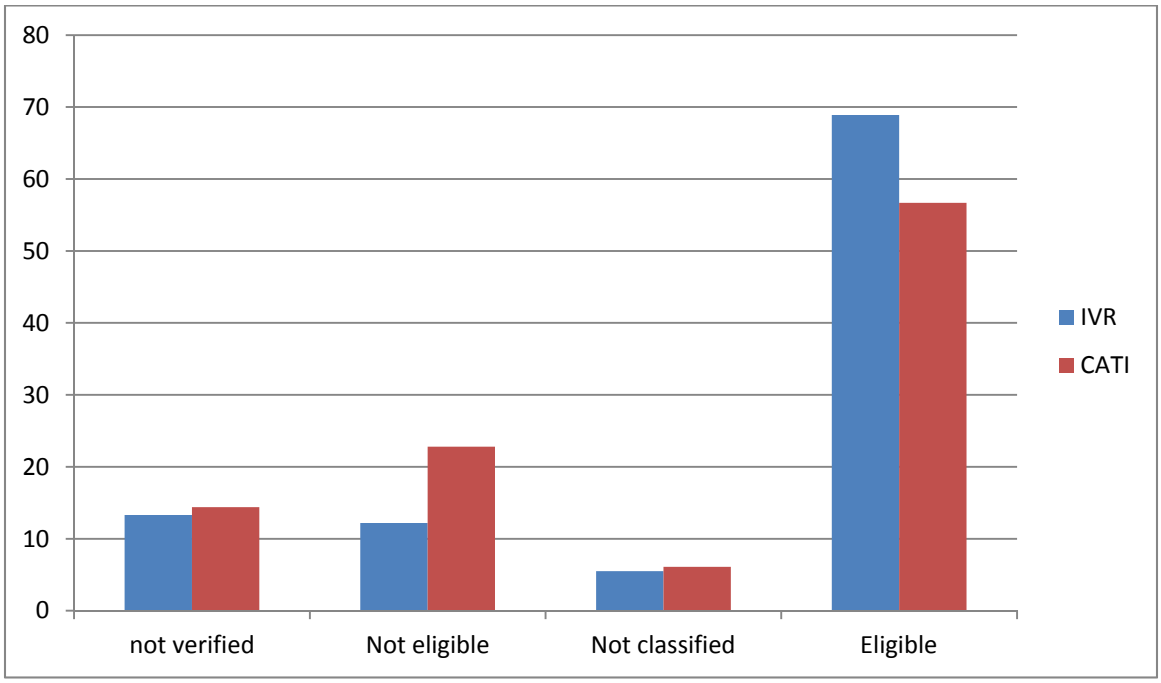
Table 7-4. Percent Distribution of Types of Crimes Pre- and Post TOC editing

Type of Crime	Pre-editing	Post-Editing
Violent Crimes	13.3%	13.0%
Rape and Sexual Assault	1.8%	1.3%
Assault	8.1%	8.2%
Robbery	2.3%	2.2%
Personal theft	1.0%	1.6%
Property Crimes	61.6%	66.1%
Burglary	22.6%	19.2%
Larceny	32.1%	40.3%
Motor Vehicle Theft	6.2%	6.9%
Not classified	25.4%	20.3%
N	1262	1262

As apparent from the above discussion, there are clearly measurement issues that are unique to adapting the IVR. These have to do with converting the questions to a yes/no format and not verifying some aspects of the events eligibility (e.g., whether the victim is a household member). But many of the above issues are also related to difficulties respondents had with negotiating the NCVS-2 questions. For example, several respondents did not interpret the location item or the evidence question as intended. Theoretically, these issues can be caught by an interviewer. We assume similar issues occur for the ongoing NCVS, since an extensive editing process is completed using the summary of the incidents. Presumably with experienced interviewers, some of the errors are mitigated, but how much is unknown.

To better understand the extent these issues are unique to the IVR, a similar review was carried out with the CATI incidents. Events were classified according to eligibility and whether the eligible event was classified correctly. We added to this mix incidents that were not verified by the respondent at the beginning of the incident form. On the CATI, this would primarily include incidents that were determined to be out of the reference period. A comparison of the classifications between the IVR and CATI are statistically different (Figure 7-2). In this case, more of the incidents on CATI were found to be not eligible. Similar percentages were either not verified or were unable to be classified.

Figure 7-2. Comparison of Quality Coding and TOC classification by Mode of Interview



There are two important caveats related to the above analyses. One is that the review of summaries was completed by one coder. While the coding was done by senior staff who were familiar with the TOC rules, this process is subject to some unreliability. Second, the coding mostly relied on the summaries at the end of the detailed incident form. Additional information was available from summaries at the screener, but these were not used. And third, the re-classification of the eligible events was not done at a 2-digit level. We did this because of time and the limited information on summaries that would prevent classification into a more detailed category. It is undoubtedly true that if a review were done at a 2-digit level, more discrepancies would be found.

7.4 Discussion

Adapting the NCVS-2 for IVR poses different challenges than NCVS-1. The NCVS-2 is structured around an interviewer working with the respondent to focus on eligible events. This requires coordination with what was reported on NCVS-1, verification that the victim is eligible, verification that the incident occurred within the reference period, and focusing the respondent on a particular incident. In addition, the interviewer administers a series of open-ended questions that are used to classify the event into a crime category. As adapted for this project, the IVR successfully carried out many of these functions. After reviewing all of the incident summaries contained at the end of the incident form, we found that 90 percent of those that were initially classified into a TOC were eligible events, while 30 percent of those not classified in a TOC should have been. This success in classification is comparable to what occurred on the CATI version of the NCVS, which was carried out by an interviewer.

However, there are issues that, if it is to be adopted for the ongoing NCVS, would need to be addressed. Some of the issues found were not a function of the IVR, but of the NCVS design itself. For example, we found that particular items on the NCVS-2 lead to measurement error and misclassification of the incident. We suspect that this type of error also occurs on the ongoing NCVS and requires data editing and coding in conjunction with the incident summaries. Nonetheless, there are also clearly issues that are unique, or more extreme, for an IVR version of the NCVS. Addressing these issues should be possible, but it will involve adapting the NCVS-2 to a self-administered mode and, specifically, to the IVR.

Review of the incident summaries found that most respondents were able to successfully navigate the transition between NCVS-1 and NCVS-2. There were some respondents where this transition

did not successfully occur for one of several reasons. Some respondents did not cleanly identify the incident that was targeted for the NCVS-2 report. These respondents had multiple incidents to report and mixed up incidents when being asked to report on these events. A second issue was reporting on events that were clearly outside the reference period. There were several instances for more serious crimes (e.g., rape) where the respondent verified the incident as occurring in the last 12 months, even though they had already stated (in the summary) that it had occurred several years ago. Finally, some respondents reported events that were not committed against anyone in the household.

These three issues can be addressed by adding more explicit verification items to either NCVS-1 or the beginning of NCVS-2. As recommended in Chapter 6, it is recommended that the month and year for each of the incidents reported on NCVS-1 be added prior to collecting any summary information. Collecting the date at this point would serve two purposes. One would be to focus the respondent on the 12 month period. While this may not prevent telescoping of events which occurred months prior to the reference period, it would prevent reporting incidents that occurred more than several years ago. The second purpose would be to provide a common reference that could be used when asking about details on NCVS-2. The IVR used the wording from the particular screening item. But this may have been inadequate when similar events, from the same screener item, were reported. With a month and year, the survey could use this in conjunction with the screener item to anchor which event is being asked about.

Events that occurred against ineligible victims, in one sense, pose less of a problem. The vast majority of these were screened out by the NCVS-2 questions. Nonetheless this error did occur and it might be best to screen these out earlier in the process, such as at the beginning of the incident form. For example, asking whether any of the victims live in the household would handle clear ineligibility for all types of crimes. For violent crimes, there is still the possibility the respondent is describing an incident against another household member. This is now handled when specifically asking about presence and the attack/threat questions, all of which refers to 'you' (e.g., "Did the offender hit you, knock you down, …" Q24). For the IVR, the respondent may ignore these references, such as one of the respondents who described a rape of her daughter. Some type of further verification that the questions are referring to the respondent may be inserted at this point.

For purposes of classifying incidents into a major crime category the IVR seemed to perform as well as the CATI version. However, this was only after review of the detailed summaries and correcting some of the classifications. Some of the errors may have been caused when converting the questions to a 'yes/no' format. For certain key questions, such as the location of the event, the

respondent has to answer the questions without knowing all of the options. For example, the questions on location started with a sequence of questions which began with:

"Did this incident happen in a vacation home, or a hotel or motel room where you were staying?"

If the respondent answered 'yes', then all of the subsequent location questions were not asked. An alternative strategy would be to provide respondents with a limited set of options (e.g., no more than three or four), which made the highest level distinctions needed. In this case, it would distinguish between 'inside home' vs. 'on your property'. Each response option would be given a number, so the respondent would simply key in the number they want to select.

The above solution could be implemented for a number of the key classification items. But it would also be desirable to reduce the number of questions. One way of achieving this is to re-structure the questions and/or use more verbatim questions with the IVR recording function. For example, there are a large number of questions on the type of property stolen. For the current IVR application, this was handled by using a series of global questions which asked about a certain type of property (e.g., cars, car parts, etc..). If the respondent had a positive response to the global, then more detail was asked. However, respondents can misunderstand these globals, for example by focusing too much on the examples. This can lead to a number of false negative and positives. It might be more efficient to ask about the value of the stolen items and record a verbatim response which describes what was stolen. This would be much easier on the respondent and likely collect more accurate information.

One limitation of the above analyses is that the primary criteria used in the evaluation of classification into a major crime category. While the TOC assignment was done at a 2-digit level, the summary information used to conduct the evaluation classified the event into one of the major crime categories. Undoubtedly there would be more discrepancies found if the more detailed categories were used in the evaluation. For example, the analysis did not distinguish between attempted and completed events, rape vs. sexual assaults and aggravated vs. simple assaults. The level of detail used was partly a function of the information contained in the summaries. If all of the summaries were used (see discussion below), it might be possible to classify to the actual TOC categories.

A second limitation is the information used to conduct the quality coding. All of the detailed incident summaries at the end of the incident form were reviewed by a single, senior, project staff member. This summary was used because it was unambiguously associated with the particular

incident. However, additional information is available from the summary given on the screener. We did a review these initial summaries for the items classified as rape and sexual assault and did find additional information about the incident. A more accurate coding would review these initial summaries to assess whether they can add to the information already collected, especially for those incidents that no useful information was found. If it is decided to move forward with the IVR, we recommend a more thorough review of this information. It could shed more light on the issues associated with the current design and suggest ways to modify it on a production basis.

Comparing Crime Rates 8

Chapter Highlights and Key Findings

- Approximately 27 percent of the entire sample and 21 percent of the ABS sample reported at least one eligible victimization. Approximately 5 percent and 3 percent of the total and ABS samples, respectively, reported a violent crime. Comparable figures for property crimes are 25 percent and 19 percent for the total and ABS samples, respectively.
- The Violent Crime victimization rate was 64 and 40 per thousand for the total and ABS samples respectively. Both of these are higher than the ongoing NCVS. The rate of rape and sexual assault is considerably higher on the IVR. The IVR property crime rate is almost twice as high when compared to the NCVS.
- The IVR prevalence and incident rates are nominally higher for those experimental conditions that have the highest response rates (mail contact mode; keypad entry mode; incentive). However, the differences are not large and only one comparison is statistically significant (property crime for the mode of contact).
- The prevalence rates for the IVR are not different from those on the CATI. This is true even when including those respondents who dropped out before completing an incident form.
- The crude victimization rates for the IVR are not different from those on the CATI. Once adjusting for missing incident forms, the IVR is nominally higher than the CATI. The rates are not statistically different, except when using the most liberal adjustment for missing data on NCVS-2. When capping the adjustment to at most six crimes per respondent or testing for the ABS sample only, none of the differences are statistically significant.
- Take-away point #1: The IVR rates are much higher than would be expected from the NCVS. It is not clear why this is the case, given the many differences between the two methodologies. One hypothesis is avidity bias. Sample in the IVR may have been more likely to respond if they had been victimized. It does not seem to be from excessive telescoping, since the differences persist even after a crude adjustment for bounding.
- Take away point #2: The experimental conditions that produced nominally higher victimization rates (mail mode of contact, incentive and keypad conditions) should be considered as a starting point if the IVR is used for the NCVS.

Chapter Highlights and Key Findings (continued)

- Take-away point #3: The proportion of persons reporting a victimization did not differ when comparing the IVR to CATI. However, there was a tendency for IVR respondents to report more victimizations per person. This may be due to undetected false positives on the IVR. Further review of the summaries on the IVR should be carried out to assess the quality of the data.

This chapter provides estimates of prevalence and victimization rates for the IVR. It addresses two questions. The first is whether the rates differ by mode of contact, mode of entry and incentive condition. The second question is whether the rates of the IVR differ from those estimated from the CATI interview.

8.1 Methods for Computing Rates

Two types of rates are computed in this chapter. One is the prevalence rate, which is the proportion of respondents that report being victimized at least once for a particular type of crime. The second is the victimization rate, which is the number of victimizations per 1000 population. The estimates concentrate on violent and property crimes, using the incident reports that have gone through the editing described in Chapter 7. Those incidents that did not have enough summary information for the quality review are based on the original TOC assignment given by the NCVS-2 responses.

IVR respondents that completed the screener and completed at least one incident form are included in the analysis. Those who completed the screener but did not fill out an incident form, when required, are excluded.

8.2 Overall Estimates and Comparisons to the NCVS

Prevalence estimates for the major types of crime are provided in Table 8-1. These estimates include all of the IVR respondents, except those in the Police sample that did not match to a telephone number. About one quarter of the entire sample reported at least one victimization. Twenty percent of the ABS sample reported at least one victimization. Table 8-2 provides the victimization rates per 1,000 population.

Table 8-1. Percent of IVR Respondents Reporting at Least One Crime by Type of Crime.*

	Total Sample		ABS Sample	
	Percent	Standard Error	Percent	Standard Error
All Crimes	27.9%	0.9%	21.3%	0.9%
Violent Crimes	5.2%	0.5%	3.2%	0.4%
Rape and Sexual Assault	0.5%	0.1%	0.5%	0.2%
Assault	3.1%	0.4%	2.0%	0.3%
Robbery	1.1%	0.2%	0.4%	0.1%
Personal theft	0.7%	0.2%	0.4%	0.1%
Property Crimes	24.6%	0.9%	19.2%	0.9%
Burglary	8.0%	0.6%	5.7%	0.5%
Theft	16.7%	0.8%	13.6%	0.8%
Motor Vehicle Theft	3.1%	0.4%	1.8%	0.3%
N	2421		1986	

*Excludes IVR sample that did not have a matching telephone number

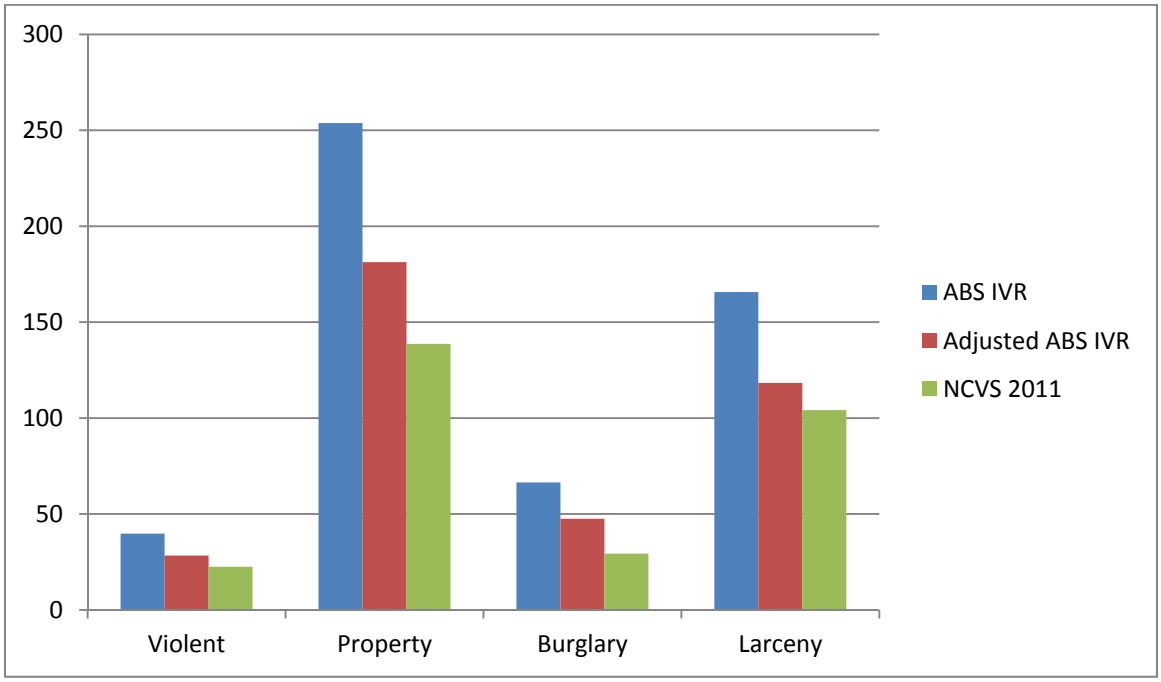
Table 8-2. Victimization rates for IVR Respondents by Type of Crime.*

	Total Sample		ABS Sample	
	Rate	Standard Error	Rate	Standard Error
All Crimes	396.5	14.7	293.6	14.2
Violent Crimes	63.6	6.0	39.8	5.3
Rape and Sexual Assault	6.6	2.0	6.5	2.2
Assault	38.0	4.6	25.2	4.2
Robbery	11.2	2.2	4.0	1.4
Personal theft	7.8	2.0	4.0	1.6
Property Crimes	332.9	13.2	253.8	12.9
Burglary	93.3	6.8	66.5	6.4
Theft	203.6	10.0	165.7	10.2
Motor Vehicle Theft	35.9	4.3	21.7	3.9
N	2421		1986	

Comparing the victimization rates to the NCVS provides a benchmark against which to judge the level of the IVR rates. It is acknowledged that there are many methodological and criminogenic reasons why the IVR and NCVS estimates will be different. Nonetheless, the NCVS can serve as an external standard against which to assess the IVR rates. For the ABS sample, the violent crime rate of 39.6 compares to 22.5 in 2011 for the NCVS. A 95 percent confidence interval around the IVR estimate is quite wide (+/- 10.6 percent), but is still considerably higher than the NCVS. The rate of Rape and Sexual assault is particularly high when compared to the .9 estimate on the NCVS. The rate for robbery and assault are comparatively lower on the IVR. In both cases, the rates are within sampling error of the NCVS estimates of 2.2 and 19.4, respectively. There is a much larger difference for the property crimes. The ABS rate of 253.8 is almost double the rate of 138.7.

One methodological feature that might be driving the IVR rates up is external telescoping. While there was a check at the beginning of the incident form to verify the timing of the incident, an explicit check on the month and year was not administered. Figure 8-1 provides a comparison of the victimization rates before and after adjusting the IVR rates after dividing by 1.4, an approximate adjustment for unbounded interviews. As can be seen, the IVR rates are still higher than the NCVS, especially for the property crimes. [5]

Figure 8-1. Comparison of IVR for ABS Sample to NCVS 2011 Before and After Adjustment for Unbounded IVR Interviews+



+ Adjustment = (IVR Rate)/1.4.

8.3 Comparison of Estimates by Mode of Contact, Incentive Treatment and Mode of entry

Tables 8-3 and 8-4 provide the prevalence and victimization rates for violent and property crimes for the three IVR-related experimental treatments. The treatments that generally have higher response rates (mail mode of contact, keypad entry, incentives) have higher crime rates. But the only statistically significant effect is for the mode of contact, where the mail mode is higher for the

[5] Methodological reasons that push the IVR rates lower are no adjustments for non-response and a longer reference period. Methodological reasons for a higher IVR estimate include that the interviews are unbounded. Criminogenic reasons the IVR estimates are higher are they only represent a metropolitan area. Reasons why they are lower are the rates exclude youth under 18 years old and the sample frame represents a more stable, older population;.

property crimes. These results hold even after running regressions which predict the prevalence or victimization rates while controlling for key household and demographic characteristics.

Table 8-3. Percent of Persons Reporting a Victimization by IVR Experimental Treatments

	Mode of Contact			
	Telephone	Standard Error	Mail	Standard Error
Violent	4.4%	1.0%	5.2%	0.5%
Property*	20.2%	1.9%	25.8%	1.0%
	Incentive			
	No	Standard Error	Yes	Standard error
Violent	4.3%	0.6%	5.7%	0.6%
Property	23.2%	1.3%	26.0%	1.2%
	Mode of Entry			
	Keypad	Standard Error	Speech	Standard Error
Violent	5.7%	0.7%	4.4%	0.6%
Property	26.0%	1.2%	23.4%	1.2%

Table 8-4. Victimization Rates per 1000 persons by IVR Experimental Treatments

	Mode of Contact			
	Telephone	Standard Error	Mail	Standard Error
Violent	57.3	14.1	65.0	6.7
Property*	254.6	26.9	350.1	14.9
	Incentive			
	No	Standard Error	Yes	Standard error
Violent	53.1	8.4	71.5	8.4
Property	310.8	19.6	349.5	17.7
	Mode of Entry			
	Keypad	Standard Error	Speech	Standard Error
Violent	71.1	8.8	55.4	8.1
Property	347.0	18.5	317.5	18.7

*Statistically significant difference at p<.0001

8.4 Comparison to CATI

One of the questions related to the IVR is whether it leads to more reports of victimizations than an interview administered instrument. With respect to prevalence rates, the two modes do not differ (Table 8-5). The prevalence rates are almost identical across the two different modes of interviewing. This rate does treats those that filled out the screener, but did not fill out any incident forms (and were supposed to) as non-respondents. If one includes these as victims, this result still holds. There is no difference between the two modes of interviews.

Table 8-5. Percent reporting at least one victimization persons by mode of interview

	Total Sample			
	CATI	Standard Error	IVR	Standard Error
Violent	4.3%	1.2%	4.6%	0.7%
Property	21.7%	2.5%	21.9%	1.3%
	ABS			
	CATI	Standard Error	IVR	Standard error
Violent	3.1%	1.1%	3.1%	0.6%
Property	15.4%	2.4%	15.6%	1.2%

A similar result is apparent for the crude victimization rates (first panel of Table 8-6). There are no differences between the two modes, with property crimes having nominally higher rates for the CATI treatment and the opposite being the case for violent crimes.

Table 8-6. Victimization Rates per 1000 persons by mode of interview and type of weight

	No weights			
	IVR	Standard Error	CATI	Standard Error
Violent	55.0	8.3	54.2	16.2
Property	290.5	18.9	314.1	40.2
	Weight =Number of Incidents Reported on the Screener+			
	IVR	Standard Error	CATI	Standard error
Violent*	114.4	23.2	54.2	16.2
Property*	497.1	67.6	314.1	40.2
	Weight = Capped Number of Incident Reports on the Screener+			
	IVR	Standard Error	CATI	Standard Error
Violent	96.0	17.7	54.2	16.2
Property	398.2	30.3	314.1	40.2

*Statistically significant difference between modes at p<.05
+ Weight = # of incidents reported on the screener/# of incident forms completed.
"Capped weight" truncates weight at 6 for those with a weight greater than 6.

As noted in previous chapters, the IVR respondents had a tendency to not fill in all of their incident forms. In addition there were respondents who reported more than three incidents, but were only asked to fill out 3 NCVS-2's. Since CATI respondents had no missing data for NCVS-2 and were not capped, comparison of the crude rates may underestimate differences between the two modes. Two adjustments were applied to estimate the effects this might have had on the differences between the IVR and CATI estimates. The first adjustment multiplied the number of completed incident forms by:

$$\text{Wgt} = (\#\text{ of crimes reported on screener})/(\#\text{ of incident forms completed})$$

For those individuals that reported three or fewer crimes on the screener and filled out all forms, this factor is '1' and the number of victimizations remains unchanged. For those that either did not fill out all forms or were capped at three, the weight is greater than 1. For example, if the respondent reported three crimes on the screener, but only filled out two incident forms, the weight is 3/2 (1.5). Once multiplying the number of incident forms completed by this factor, the number of crimes for that individual will be '3' (1.5 * 2 = 3). This adjustment accounts for both missing incident forms, as well as the capping procedure. A second weight was calculated that truncated the first weight to 6. There were respondents who reported many crimes on the screener (e.g., >50). The capping reduces the influence of these outlier cases.

This weight does not distinguish between different types of crimes. A more refined adjustment might calibrate the weight by the screener items (e.g., personal vs. property crimes). But this procedure does provide a sense of the effects of missing incident forms, which only affect the IVR.[6] When applying the weights, the standard errors go up significantly for the IVR sample. This reflects the influence of those respondents who reported a large number of incidents on the screener.

The direction of the differences change when using the full weight. For both types of crimes, the IVR produces significantly higher rates of crime ($p<.05$). These differences are not statistically significant when using the capped weight, but they are still in the same direction.

Some of the difference in victimization rates are heavily influenced by respondents in the police sample, who disproportionately reported multiple crimes on the screener. Once taking these out (Table 8-7), the IVR yields more reports of crime than the CATI. But the differences between the IVR and CATI are smaller and are no longer significant even for the fully weighted estimates.

Table 8-7. Victimization Rates per 1000 persons for ABS sample by mode of interview and type of weight

	No weights			
	IVR	Standard Error	CATI	Standard Error
Violent	37.6	7.9	35.1	13.7
Property	195.1	16.9	219.3	37.6

[6] Note this procedure does not affect the prevalence rates, which count those that filled out at least one incident form.

Table 8-7. Victimization Rates per 1000 persons for ABS sample by mode of interview and type of weight (Continued)

	Weight =Number of Incident Reports on the Screener+			
	IVR	Standard Error	CATI	Standard error
Violent	55.8	14.9	35.1	13.7
Property	297.9	66.6	219.3	37.6
	Weight = Capped Number of Incident Reports on the Screener+			
	IVR	Standard Error	CATI	Standard Error
Violent	52.3	12.7	35.1	13.7
Property	236.8	23.2	219.3	37.6

+ Weight = 1 for those who did not report a victimization and those that filled out all incident forms.
=# of expected reports/# of incident forms filled out. "Capped weight" truncates this at 6 for those with a weight greater than 6.

To check whether the above results were influenced by differential non-response, regressions were estimated which predicted the prevalence and victimization rates once controlling for demographic and household characteristics. The above results hold for these regressions in all cases.

8.5 Discussion

The rates from the NCVS IVR are high relative to those produced by the NCVS. There are many reasons why the two estimates might differ. For example, the IVR is an unbounded interview, which would lead to higher rates. A difference that pushes the IVR rates down is the absence of a non-response adjustment for age. Age is an important characteristic that is not accounted for in this comparison. The response rate for young people for the IVR was low, as noted in prior chapters. Young people also have much higher rates of victimization. An adjustment for age would significantly increase the rates shown in the above analysis. Further research into the reasons for the differences should be explored as the NCVS is redesigned in the future.

The different experimental treatments that had the highest response rates (i.e., mail mode of contact; incentive; keypad) also had nominally higher rates of victimization. However these differences were relatively small and only one was statistically significant (property crime for mode of contact). Some of this difference may be attributable to a larger number of missing incident forms in the low response rate conditions. As shown in previous chapters, those in the speech mode and non-incentive conditions tended to have a smaller proportion of incident forms filled out. Nonetheless, this result supports the possible adoption of these treatments if an IVR option is used as part of the NCVS.

The IVR did not differ from the CATI with respect to prevalence rates. The rates were almost identical for both types of crimes. The victimization rates also did not significantly differ when relying solely on the incident forms that were completed. After adjusting for missing incident forms and the capping procedure, the IVR victimization rates were consistently higher than the CATI. These differences were only statistically significant when weighting to fully account for all screener reports of incidents. If the weight is capped at six or if one just includes the ABS sample, none of the differences are statistically significant.

The higher number of victimizations reported in the IVR condition, may partly reflect the unverified nature of these reports. Whether it was because the incident was capped or the respondent dropped out of the survey, these incidents did not go through NCVS-2 and, therefore, may not represent an eligible crime. If the IVR is to be adopted, we recommend a review of the incident summaries on NCVS-1 to further assess the quality of the data reported on the screener.

Costs of the IVR 9

Chapter Highlights and Key Findings

- The cost of the IVR Mail procedure ranges from $21/complete to $37/complete, depending on various assumptions about the design. This compares to the cost of a CATI survey, which ranges from $80/complete to $120/complete
- Larger samples significantly reduce the cost of the survey because it is possible to amortize the fixed management and set-up costs. A promised $10 incentive increases the cost by approximately $4/complete.
- Take-away point #1. The IVR Mail procedure is significant less expensive than a CATI survey. If possible to manage, a local area agency could achieve double or triple the sample sizes on fixed budget that is allocated to a CATI survey.
- Take-away point #2. For the ongoing NCVS, the option of using the IVR to administer the NCVS-1 could result in significant savings. In one hypothetical example, it is calculated it could save $600,000 out of a budget of $4 million. More specific cost modeling should be completed when considering moving forward with IVR in this capacity.

One of the advantages of the IVR is that it is a relatively inexpensive method to implement, once the survey is designed and developed. In this chapter we summarize the costs of implementing the IVR Mail design used on the current project.

9.1 Cost Assumptions

The estimates below include the cost of the labor and materials associated with implementing the IVR Mail survey. Table 9-1 provides a list of the different components that were included in the calculation.

The estimates used a 2.5 month data collection period. This assumes that the organization conducting the survey has the computer server capacity to handle all calls that might come in during the survey period. If capacity is not large enough, then the survey period would have to be lengthened because of the need to stagger the mailings to control the number of incoming calls.

Costs were estimated assuming the response rates observed in the study. Two separate costs were derived. One is for an ABS sample that matched addresses to telephone numbers. This is identical to what was used for this project. The response rates assumed for this estimate were based on the rates for the ABS sample discussed in Chapter 3. The second set of estimates is for a design which mailed to all households, without matching to telephone numbers. The response rates assumed for the cost estimates were based on two sets of rates. One was the unmatched police sample discussed in Chapter 3. The second were those published in Montiquila, et al.(2013).

Figure 9-1. Labor and Other Direct Cost Components for estimates of IVR Mail Procedure

Labor
Project Manager
IVR Programmer
Mailing staff
Other Direct Costs
Envelopes, stationary
Inserts
Printing
Incentives
Time on computer servers
Telephone connect time

9.2 Costs

Table 9-2 provides the costs of conducting the Mail IVR under different scenarios related to the size of the sample and whether or not a promised incentive is used. These figures are provided as the cost per completed survey.

Table 9-2. Estimated Cost Per Completed Interview

	Matched Sample		Total Population	
	Promised $10?		Promised $10?	
N	No	Yes	No	Yes
10,000	$33	$37	$37	$41
50,000	$21	$26	$24	$28

The cost of the IVR Mail procedure ranges from a low of $21 per complete to $37 a complete. Most of these costs are those related to the labor and materials associated with the mailings. The costs go down as the sample size goes from 10,000 to 50,000. This is expected because the fixed

costs associated with the survey, such as labor to manage and monitor the survey, does not significantly increase with sample size. This leads to greater amortization of these fixed costs as the sample size gets larger.

The costs go up marginally when surveying the total population, rather than just those where a matched phone number can be found. When both matched and unmatched addresses are included, the response rate goes down, which drives the costs of the mailings up.

The promised incentive adds approximately $4 for each completed interview. It does not equal $10 a complete because the survey gains efficiency when the response rate goes up as a result of the incentive.

Based on industry rates, average CATI costs for a survey of the complexity of the NCS range from $80 to $120 a completed interview. The range reflects different procedures that might be used to contact respondents. As expected, the IVR is considerably less expensive than CATI.

9.3 Discussion

The cost of the IVR Mail procedure is considerably lower than a survey administered with a CATI. For a local survey, this is a significant savings and would allow agencies to double or perhaps even triple sample sizes for a fixed budget.

It is difficult to project exacted cost savings if an IVR were to be used as a follow-up for the NCVS. For example, the costs of the IVR might be considerably lower because there would not be a need to send out multiple mailings to each respondent. Perhaps a single letter reminding everyone to use the IVR might be sufficient. On the other hand, one would expect there to be some interviewer or home office involvement with monitoring responses and sending out reminders when respondents do not use the IVR.

For purposes of discussion, calculations were completed that assume that the costs of the IVR stay approximately the same as shown in Table 9-2 when following up NCVS respondents. It was also assumed that the design administers the NCVS-1 using the IVR at the second time in sample. If it costs approximately $80 for an interviewer to conduct a telephone interview with respondents at the second time in sample, then there is a cost savings of between $50 - $60 for each individual who

completes an IVR interview and does not need any follow-up (e.g., , because no victimization was reported).

This translates into significant savings. If one assumes that 50,000 respondents are eligible for an IVR interview, the baseline cost of conducting telephone follow-ups would be approximately $4 million ($80 x 50,000). If one further assumes that 30% of the sample use the IVR and that 80% do not report a victimization, this would save approximately $600,000, assuming that approximately $50 is saved for each IVR interview (.3 x .8 x 50,000 x $50). The 30% estimate is based on the 30% response rate achieved with the IVR Mail procedure. The 80% is based on the prevalence rate measured on the IVR screener.[7]

This is a purely hypothetical scenario, since it is not based on actual cost data for either the NCVS or what an IVR would cost in the context of the NCVS. Nonetheless, this example provides a general idea of how an IVR might lead to significant savings within the context of the NCVS. More concrete cost estimates can be provided by obtaining information on the cost of current NCVS interviews and more detailed cost modeling of different designs that might be used if the IVR were incorporated within the NCVS.

[7] The prevalence rate shown in chapter 7 was 30% for a 12 month reference period. The 20% used in this calculation adjusts this down to reflect a 6 month reference period.

10 Discussion and Recommendations

This chapter summarizes the answers to the research questions discussed in Chapter 1 and makes recommendations on areas of future research if an IVR is considered for implementation on the NCVS.

10.1 Research Questions

There were three major research questions that motivated this project: 1) what are the response rates for the IVR?, 2) Can the IVR be adapted for use on the NCVS and 3) What are the victimization rates for the IVR. Chart 9-1 provides a summary of the results from the experiments, how they relate to each of these questions and the general recommendations based on these results.

Research Questions	Answer to Question and recommendation
OMB 2. What are the response rates with IVR? 1. What are the response rates using inbound and outbound contact methodologies (mail contact vs. telephone contact)? 2. What is the composition (e.g., socio-demographics) of those that respond to the IVR and how does this differ by mode of initial contact? **3.** Is it possible to effectively encourage sampled households to complete the IVR interview a) Do nominal response rates increase respondents' willingness to utilize IVR in a self-administered interview b) Does an insert encourage sampled households to respond to the IVR interview	1. Response Rates for inbound (mail) are around 23 percent for the NCVS-1 and 21 percent for NCVS-2. They are 5 percent lower for outbound Keypad respondents completed at a higher rate than Speech About 30 percent of those with >=1 victimization do not fill out all required forms. 2. Both contact methods result in under representation of young people, low education and Hispanics. Mail contact mode does somewhat better with respect to getting younger respondents. 3. It is possible to effectively encourage respondents using an incentive and an insert. A $10 incentive for the mail contact method increased response rates by 6 points. The incentives increased the percentage of respondents completing all NCVS-2s by 10 percentage points. An insert with an incentive adds approximately 3 percentage points to the response rate. There was no effect of the insert in the no incentive condition. <u>Recommendation</u>: .From a response rate perspective, it should be possible to incorporate the IVR on the NCVS <u>For ongoing NCVS</u>. Given equivalent response rates with CATI, it seems reasonable that it would be accepted by many respondents as a follow-up to T2 – T7 interviews. On response rates, NCVS-1 is more acceptable than NCVS-2. <u>For local area surveys</u>. If comparison is to CATI, NCVS-1 response rates are equivalent for the IVR. NCVS-2 rates are lower.

Research Questions	Answer to Question and recommendation
Can the NCVS questionnaire be adapted for IVR administration? 1. Are respondents able to complete the two major components of the NCVS (NCVS-1, NCVS-2)? 2. Can the IVR handle tasks normally done by the interviewer, such unduplicating incidents and transitioning to the NCVS-2 3. What is the quality of the incident data from the IVR? 4. Are there differences in respondent acceptance between speech IVR and touchtone data entry (Keypad)?	Yes, it can be adapted. NCVS-1 is closer than NCVS-2 1. Respondents did complete the NCVS-1 at rates equivalent to the CATI. Very few dropouts. There many more dropouts when filling out the NCVS-2 (see response rate section) 2. The IVR can handle these tasks for the NCVS-1. For the NCVS-2, it works for most respondents. But there are still issues that need to be addressed related to detecting ineligible victims, incidents that are out of the reference period and making sure the respondent understands which incident they should describe on NCVS-2. 3. The quality of the data for the NCVS-1 is good. There is a tendency for IVR respondents to report more victimizations, perhaps some of which are false positives. NCVS-2 worked for most respondents. But there were quality issues related to the structure and wording used to adapt the IVR. 4. Keypad respondents were much more likely to complete all phases of the survey. They experienced many fewer entry errors when compared to speech respondents. <u>Recommendation</u>: IVR can be used to screen respondents for either the ongoing NCVS or a local area survey. The NCVS-2 could also be used, but more research should be completed to tailor the instrument for the NCVS. Keypad entry should be used. Speech could be offered as an optional mode, but total reliance on speech needs further advances in the technology. If IVR is used by local areas, specialized technical assistance will be required to develop and manage the work.

Research Questions	Answer to Question and recommendation
What are the Victimization Rates for the IVR? 1. Does IVR lead to different victimization rates from a telephone interview? 2. Is there a difference in victimization rates for Speech and Keypad modes of entry? 3. Is there a difference in victimization rates by method of contact or incentive?	Rates were generally higher than the ongoing NCVS, even after adjusting for the unbounded nature of the procedure. 1. Without adjustment for missing data, there was not a statistically significant difference in the prevalence or victimization rates between the IVR and CATI. This result persisted even after controlling for demographics of respondents. The victimization rates were crudely adjusted for non-response (failure to fill out all victimization forms) and the capping procedure. Once adjusted, the victimization rates for the IVR were nominally higher than CATI. Results were statistically different for the total sample, but not for the ABS sample alone. 2. There were no differences in prevalence and victimization rates between different modes of entry. 3. There were no differences in prevalence and victimization rates for the incentive and mode of contact experimental conditions for any of the crimes. The one exception was there was a significant difference for property crimes for mode of contact. Recommendation: The IVR produced similar prevalence rates as the CATI, but nominally higher per person rates of victimization, once accounting for missing data. Future research should examine this potential mode effect more closely by assessing the quality of the reports of victimizations from the NCVS-2.

10.1.1 What are the Response Rates for the IVR?

Findings and Recommendations

- From a response rate perspective, the IVR will perform equivalent to a CATI interview for NCVS-1. Use for NCVS-2 is also possible, but likely involves more dropouts.
- IVR Mail, keypad, an incentive and insert will maximize response rates.
- Use for local area surveys could be considered when CATI is the alternative.
- Use for ongoing NCVS should be considered as follow-up at 2nd through 7th times in sample

The experiment examined two questions related to response rates. One was to compare response rates by different experimental conditions. The second was to test two methods to enhance the IVR response rate.

Comaprative response rates. This experiment provided evidence that with respect to response rate for NCVS-1 the IVR Mail methodology is equivalent to CATI. The IVR Telephone methodology is less viable when considering the response rates. There were a significant number of individuals who abandoned the interview during the transfer between the interviewer and the IVR. The keypad respondents had higher response rates than the speech respondents. With respect to respondent characteristics, there were very large differences between all of the experimental groups and the general population. When comparing to ACS data, all of the groups underrepresented young people, those with a lower education and Hispanics.

The IVR response rates drop once considering completion of NCVS-2. Overall, the IVR-Mail and IVR-Telephone rates for the ABS sample drop 2 – 3 percentage points. While this is a relatively small drop in the overall rate, this hides the fact that a significant number of individuals did not complete all of the NCVS-2 forms that were expected. Overall, approximately 30 percent of the respondents did not fill out all of the required incident forms. This compares to a 100 percent completion rate for the CATI.

Enhancements to the response rate. The results support the use of a promised incentive of $10 to encourage use for the IVR Mail condition. When paired with the insert, the response rate went up by approximately 10 percentage points. The promised incentive did not work as well for the IVR Telephone group. The incentive also showed some evidence that data quality was enhanced with more respondents filling out all detailed incident forms and less missing data on income. However,

there was no evidence that the enhanced response rate significantly improved representation of demographic groups.

Implications. From a response rate perspective of completing NCVS-1, the IVR Mail condition is at least equivalent to a CATI interview. If used in conjunction with a keypad entry mode, an incentive and an insert, it performs significantly better. Completing the NCVS-2 is more problematic. A significant number of respondents did complete multiple forms. However, there was a significant drop-off in response, with 30 percent of respondents not completing all of the forms required of them. These results suggest that the IVR Mail procedure is promising as either a follow-up to NCVS respondents (e.g., at T2 – T7) or as a way to screen respondents for a local survey.

10.1.2 Can the NCVS be Adapted for IVR Administration?

Findings and Recommendations

- The keypad mode had significantly fewer entry errors. The accuracy of the data-entry was also better than the speech. The keypad mode of entry should be used for any future IVR application.
- The IVR can be used for NCVS-1, especially with a few more adaptations (e.g., verifying month/year; verifying the eligibility of the victim).
- For NCVS-2, the IVR performed reasonably well when classifying incidents into the major crime categories. The results suggest that it would be feasible to use the IVR for NCVS-2, although more work is needed to improve data quality.
- To improve quality on NCVS-2, several changes should be made relative to the IVR application used on this project: 1) re-structure questions for a keypad mode and 2) use more open-ended verbatim items.

This question involved examining several aspects of data quality related to the IVR method across NCVS-1 and NCVS-2. The first was to examine the overall usability of the IVR. The second was to examine how well the IVR completed tasks normally done by the interviewer for NCVS-1. The third aspect was to examine how well these tasks were completed for NCVS-2.

Usability. The average time to complete the NCVS-1 screener ranged from seven minutes to 10 minutes, depending on whether a victimization was reported. Once a victimization was reported, the time to complete goes up to between 16 and 24 minutes, depending on the number of NCVS-2s

filled out. The speech respondents had the highest error rates, with the driving problem being the system failing to recognize their responses. On NCVS-1, 76 percent of the speech respondents had at least one question where the system did not recognize their response. There were also several problems with the design of selected IVR questions, including income and selected 'mark all that apply' items. Those who had at least one victimization were more likely to report being less satisfied with the questionnaire, confused or frustrated and that the survey was too long. Finally, there were instances where the IVR system, primarily for the speech respondents, did not correctly code the data.

Adapting the NCVS-1. Adapting the IVR for NCVS-1 was reasonably successful at collecting reports of victimizations. The procedures in place were successful at detecting duplicate incidents. Similarly, respondents did take advantage of the verification procedure at the beginning of the incident form to correct reports on the screener.

With respect to actually reporting victimizations, there was a small, but significant difference between the two modes of contact, with the IVR Mail having a higher proportion reporting victimizations than the IVR Telephone. This result holds up after controlling for differences in household and demographic characteristics. When compared to the CATI, the IVR had a nominally higher proportion of persons reporting a victimization, the differences were not statistically significant.

The number of victimizations reported for the IVR was high. For the ABS sample, 40 percent of those reporting a victimization had at least two incidents. The number reported on the IVR was significantly higher than the CATI. While there was a tendency for IVR respondents to reduce these numbers at the verification at the beginning of the NCVS-2, this only affected about 10 percent of the respondents reporting a victimization.

Adapting the NCVS-2. Adapting the NCVS-2 for IVR poses different challenges than NCVS-1. As adapted for this project, the IVR successfully carried out many of the required functions. To assess the quality of the NCVS-2 data, all summaries contained at the end of the incident form were reviewed and compared to the initial TOC assignment. This review found that most respondents were able to successfully navigate the transition between NCVS-1 and NCVS-2. There were some respondents where this transition did not successfully occur for one of several reasons: 1) some respondents did not cleanly identify the incident that was targeted for the NCVS-2 report, 2) some respondents reported on events that were clearly outside the reference period, and 3) some respondents reported events that were not committed against an ineligible victim.

For purposes of classifying incidents into a major crime category the IVR seemed to perform as well as the CATI version. About one-third of events classified into a TOC were in the wrong category. Once reviewing these incidents, it was found that most of the error was misclassification within the major crime groups of violent and property crimes. For example, a significant number of incidents initially classified as a burglary were actually thefts or motor vehicle thefts. The opposite error also occurred. A large proportion of these errors occurred because respondents did not report a theft when asked. A second type of error occurred because respondents mis-interpreted key questions needed in the TOC algorithm. For example some respondents reported events in the yard as being 'in home', which lead to incorrectly assigning a burglary TOC. Some of this error seemed to be a result of breaking up the questions into a series of 'yes/no' items. As a result, respondents did not hear all of the response alternatives.

Implications. The evidence on overall usability and adapting the instrument provide further evidence that NCVS-1 can be used for the NCVS or a local area survey. We are recommending using a keypad mode of entry to minimize entry errors and levels of frustration on the part of the respondent. The design might also consider giving respondents a choice on which mode to use, however the design of the survey should be driven by the keypad entry mode. It should be possible to incorporate speech for the IVR as the technology in speech recognition advances.

The evidence seemed to indicate a tendency of IVR respondents to report more crimes per person. The prevalence rates were very similar between CATI and IVR, but the number of victimizations reported was higher for IVR. This raises the possibility that the NCVS-1 reports from the IVR may contain ineligible events. One change that should be considered is to have a more explicit confirmation of the month/year of each incident reported. It would be preferable to collect the month and year for each incident at the time the incident is initially reported on NCVS-1. This will force the respondent to think about the timing of the events relative to the reference period. Additional verification at the screener will increase the chances that respondents will report on an eligible event. For example, verifying the identity of the victim of the incident would prevent respondents from reporting other types of ineligible events.

Use of the IVR for NCVS-2 will involve adapting the instrument more explicitly to a keypad mode. The IVR for this project was designed to accommodate both keypad and speech. All questions were converted to yes/no items to ensure the speech would have minimal problems recognizing utterances. This format was not ideal to adapting critical open-ended NCVS-2 items, such as location and type of items stolen. One adaptation will be to translate the current open-ended NCVS-2 items to multiple choice questions, which can be answered with numbers on the keypad.

This would allow respondents to hear all of the response categories before answering. A second adaptation would be to use open-ended, verbatim, questions whenever possible. For example, asking about which items were stolen. Use of this type of question would reduce the overall burden of the instrument.

10.1.3 What are the Victimization Rates for the IVR?

Findings and Recommendations

- The IVR produced rates of victimization much higher than the NCVS. It is unclear what is behind these differences, but this difference is also apparent for the CATI application and several other studies that have tried to replicate the NCVS methodology.
- The IVR had nominally higher victimization rates for those experimental conditions that had the highest response rates (i.e., mail contact, incentive and keypad). However, there were no significant differences across the different experimental conditions, except for property crimes being higher in the IVR Mail contact mode.
- The prevalence rates between IVR and CATI were almost identical. The victimization rates for IVR were higher once accounting for missing NCVS-2's and the capping procedure. This suggests the IVR produces higher reports of multiple victimizations and, perhaps, more false positives.

This question involved estimating the prevalence and victimization rates with the IVR and comparing these rates to the NCVS, across the experimental conditions and to the CATI.

Comparison to the NCVS. The rates from the IVR are high relative to those produced by the NCVS. The biggest difference is for property crimes, where the IVR rate is almost twice as high as the NCVS. This difference persists even after making a crude adjustment for bounding on the IVR. Given the large number of differences between the two surveys, it is impossible to pinpoint why they differ. However, these results are consistent with several other surveys that have applied the NCVS methodology (Westat, 2013; Biderman, et al., 1985).

Comparison of experimental conditions The different experimental treatments that had the highest response rates (ie., mail mode of contact; incentive; keypad) also had nominally higher rates of victimization. However these differences were relatively small and only one was statistically significant (property crime for mode of contact). Some of this difference may be attributable to a larger number of missing incident forms in the low response rate conditions. As discussed

elsewhere, the low response rate conditions tended to have a smaller proportion of expected incident forms filled out.

Comparison to CATI The IVR did not differ from the CATI with respect to prevalence rates. The rates were almost identical for violent and property crimes. The victimization rates also did not significantly differ when relying solely on the incident forms that were completed. After adjusting for missing incident forms and the capping procedure, the IVR victimization rates were consistently higher than the CATI. These differences were only statistically significant when weighting to fully account for all screener reports of incidents. If the weight is capped at six or if one just includes the ABS sample, none of the differences are statistically significant.

Implications. The high rates of the IVR relative to the NCVS cannot be easily explained. While there are many differences in methodology between the two, many of these differences would suggest the IVR would be lower than the NCVS. For example, the rates were not adjusted for non-response. The IVR is based on a sample that likely has lower victimization rates than the general population (i.e., those with matched phone numbers). The design only interviewed one person in the household, so reports of theft, which are most affected by multiple respondents, should be underreported. But the biggest differences between the NCVS and the IVR are for these crimes. This large difference is not solely for IVR, the CATI estimates were similarly different. One possible explanation is linked to a form of non-response bias which predicts that those cooperating on the IVR are self-selecting into the study because they have been victimized. We could not test this hypothesis. We only note that one other study that replicated the NCVS, which had a much higher response rate, found similar differences (Biderman, et al., 1985).

There was inconsistent evidence of a mode effect when comparing the IVR to CATI. There was no evidence when examining prevalence rates, while there was a tendency for the IVR to yield higher reports of multiple victimizations. This may partly reflect the unverified nature of the IVR reports. Whether it was because the incident was capped or the respondent dropped out of the survey, these incidents did not go through NCVS-2 and, therefore, may disproportionately represent an eligible crime. If the NCVS-2 is to be adopted, we recommend further review of the incident summaries taken at the screener in order to assess the quality of the data.

10.2 General Recommendations and Next Steps

The two proposed uses of the IVR are: 1) to follow-up respondents after the initial household visit and/or 2) to conduct a local area survey. In both cases, the IVR could be used in two different ways. One would be administer NCVS-1 to screen for victims. The second is to administer both NCVS-1 and NCVS-2 to estimate victimization rates.

10.2.1 Use of IVR as a Follow-Up to the Ongoing NCVS

Under this scenario, respondents interviewed at the first time-in-sample would be asked to respond at the next time-in-sample using the IVR or are given a choice between the IVR and a Web survey. Without any data on the performance of a web application, it is difficult to know how it compares to the IVR. But clearly the Web should be considered along with the IVR, if not before, as a viable self-administered mode for the NCVS.

Respondents should be offered a small incentive, such as $10, to complete the survey using the IVR (or web). If NCVS-1 is the only portion administered via the IVR, respondents who report a victimization would be followed up with a live interviewer by telephone. This would not be necessary if the NCVS-2 was also administered.

This project provides evidence that this could lead to a significant reduction in the number of interviewer administered surveys. With a 12 month, unbounded, reference period, about 30 percent of the respondents reported a victimization. Using this as a baseline, the above design could reduce the workload of the interviewer by 70 percent for all of those that complete the NCVS-1 by IVR. If 30 percent of respondents actually use this option, the number of interviews that are administered would be reduced by around 20 percent (70 percent x 30 percent). For those the interviewer does follow-up, only NCVS-2 would have to be administered.[8]

If the NCVS-2 is also administered by IVR, then no follow-up would be required, at least for those that fill out all required incident forms.

[8]One would expect that the percentage reporting a victimization will go down with a six month reference period and at later times in sample. It is unclear whether 30% is a realistic estimate of the proportion of respondents choosing to complete the IVR.

Under both scenarios, it is recommended there be a check, perhaps by the home office or by the interviewer, to see if data are provided during the specified time period. Those that haven't responded would be prompted, perhaps reminding them of the incentive, before the interviewer follows up to complete the interview.

The results of this project support both of the above applications. The application for the NCVS-1 is closer to being ready than the NCVS-2. This project found more problems with measurement and response rates on the NCVS-2. This is consistent with prior research, which has found that the most appropriate survey applications for an IVR are relatively short, simple, surveys. The NCVS-1 is relatively short and simple to complete. Filling out the NCVS-2 added at least six minutes to the survey task (e.g., from 10 minutes to 16+ minutes). The response task for the NCVS-2 is also more complex. Respondents are asked to describe the event, sometimes in their own words.

On the other hand, the results from this project found that many respondents were willing to complete most incident forms (70 percent completed all forms). On the ongoing NCVS, this non-response may not pose an insurmountable problem. For example, it would be possible to conduct non-response follow-up for those who do not complete all forms. Reducing measurement error may pose more challenges. This project found that about 30 percent of the incidents had to be re-classified into a different category after comparing the summaries to the assigned TOC. It might be possible to reduce this error through redesign of the IVR version of the NCVS-2 and, more generally, the content of the NCVS.

To further assess the feasibility of the IVR for the NCVS, additional research should be completed. First, it would be useful to further explore the effects of mode on reports of victimization. For both the NCVS-1 and NCVS-2, there seemed to be a tendency for the IVR respondents to report more multiple victimizations when compared to the CATI respondents. One suspicion is these are reports of ineligible events which were not verified by going through NCVS-2. The IVR may have also lead to better recall and reporting of events that are under-reported in an interviewer-administered survey. More research on these two possibilities should be conducted. A relatively low cost investment would be to review the summaries from this project to further describe the types of events that were reported on the IVR screener. These reviews would checks on whether the events are eligible with respect to the reference period and the victim. The present project only reviewed the summaries at the end of NCVS-2. Consequently, summaries were not examined for those incidents reported on the screener but no incident form was completed, either because the respondent dropped out or more than three incidents were reported. But even for those incidents

which had an NCVS-2 summary reviewed, there may be more information on the NCVS-1 summary that sheds more light on the incident.

A second line of research would be to experiment with adapting the NCVS more specifically for this mode of interviewing. For NCVS-1, this involves integrating more verification items into the screening process. Specific checks should be administered on the month and year of each incident and the eligibility of each victim (e.g., property belonging to the household; respondent is victim of violent crime). For NCVS-2, the adaptation involves designing multiple choice items and the use of more open-ended verbatim questions whenever possible.

More generally, when considering any self-administered mode within the ongoing NCVS, some investigation of the mode effects needs to be done within the context of the NCVS. The current study was a comparison with a CATI administered, one-time survey. How CATI and IVR compare when the interview is conducted by an experienced NCVS interviewer, either over the telephone or in-person, should be examined. Similarly, if other self-administered modes are being considered, such as the Web or audio computer assisted self interviews (ACASI), then comparisons to these modes should also be completed.

The mode effects should be examined when the survey is used as a follow-up of respondents who have already been interviewed. The present study was conducted as a one-contact design. NCVS respondents generally report fewer incidents at higher times-in-sample. While the reasons for this decrease are unclear, some of the change is due to respondents learning more about the survey and what it covers. Interviewers may also change their behavior at later times-in-sample. The assessment of mode effects in the present study does not take this into account.

In conjunction with this methodological research, more specific cost models should be developed to better understand the cost savings that might be realized with different types of designs. The hypothetical cost savings discussed in this report suggest that significant savings can be achieved. But these estimates were not based on actual costs of the ongoing NCVS nor did it put the IVR within a specific set of design parameters. With more information on each of these, it should be possible to get a better sense of the cost benefits of the IVR for the NCVS.

10.2.2 Use of IVR for Local Surveys

The main advantage of the IVR for a local area is that it is relatively inexpensive to administer. Relative to a web or CATI survey, the IVR does require more specialized technical expertise. If local area agencies were to use the IVR, they would need fairly detailed programming specifications to maximize data quality, as well as control development costs. Given this, it is not clear whether an IVR would be less expensive for a local agency than other types of surveys that can draw on more familiar survey procedures, such as a CATI survey. If an IVR were to be a viable option for a local area, it would be beneficial for BJS to provide the programming specifications for the basic survey. The local area could modify these specifications to meet their own needs.

There are unique tasks for which the IVR could be inexpensively used. For example, the IVR Mail procedure could administer the NCVS-1 to find individuals who have been victimized. These individuals could be followed up using interviewing methods to ask more detailed questions. This might particularly useful if the goal is to find individuals who have been a victim of particular types of rare crimes (e.g., domestic violence; burglaries). A mail survey offers a similar capability, except it is not able to incorporate complex skip patterns or procedures. For example, it would be difficult to un-duplicate events on a mail survey as on the IVR. A web survey offers similar advantages to the IVR.

This study found the victimization rates for the IVR were somewhat higher than for a CATI application. If BJS is interested in offering IVR to local agencies, the research noted above should be carried out. In particular, the design of the IVR should be adapted to a keypad mode of entry and more research should be conducted on why IVR respondents seemed to report more victimizations when compared to respondents to the CATI.

10.3 Limitations of the Study

The limitations of this study are linked to the sample frame used, the relatively small sample sizes for the telephone modes, the measures of data quality for the NCVS-2 and the length of NCVS-2.

The study used an address based sample (ABS) which were drawn from the general population and from central city police departments in two metropolitan areas. The addresses were matched to find a telephone number. With a minor exception, the final sample only included those addresses that

had a matched telephone number. This was done to facilitate comparisons between the IVR Mail treatments to those that were contacted by telephone (IVR Telephone, CATI). Consequently the results cannot be directly generalized to the entire population of interest to the NCVS. Residents of households with an address that can be matched to a phone number tend to be more cooperative on surveys. For example, in the one portion of the sample that was not matched to a phone number, the IVR response rate was 6 to 7 percentage points below the sample with a matched phone number (see also Montiquila, et al., 2013). It is not clear how the victimization rates might be affected. Those with an unmatched number do tend to be younger and more geographically mobile. Both of these characteristics are positively correlated with victimization.

The comparisons across the different modes of contact and interviewing are based on relatively small sample sizes. For example, the CATI completed approximately 280 interviews, while the IVR Telephone interviews completed approximately twice as many. For survey-related outcomes, such as prevalance and victimization rates, statistical tests comparing these to the IVR Mail treatments have relatively low power. For example, the comparisons between CATI and IVR Mail prevalence rates had about 80 percent power to detect an 8 percentage point difference between the two modes. For a base prevalence rate of 20 percent, this represents about a 40 percent difference. This relatively low power biases the results against finding significant mode differences that were smaller than this.

The third limitation of the study is related the review of the summaries taken from the end of the incident form. This review was completed to assess the quality of the data on the IVR version of the NCVS-2. The results of this analysis found some measurement error on the IVR version of the NCVS-2 with respect to classifying events with the TOC algorithm. The review of summaries were all done by one individual who was experienced with the TOC assignments. The coding was limited by the information that was available in the summaries. It was also limited by the use of a single coder, without any independent coding of a subset of the summaries. As recommended above, future research into the quality of the IVR data should review all of the summaries available for each incident (ie from both NCVS-1 and NCVS-2). It may also be beneficial to have several different individuals, who are familiar with the TOC algorithm and classification rules, code at least a subset of cases in order to estimate a measure of reliability for the codes.

Finally, the IVR on this study modified the NCVS design in ways to make it more compatible for this mode. This involved shortening the survey and making the questions simpler and easy to provide a response. Wrapped in these adaptations is shortening the NCVS-2 to those questions that were needed to classify the incident. Some additional questions were retained, such as whether the

police were involved. But overall, most other non-classification related questions were removed. The results reported above are for this shortened version of the NCVS and could be different if a longer version of the survey were used.

References R

Biderman, A.D., Cantor, D., Lynch, J.P., and Martin, E. (1985). *Final report of the National Crime Survey redesign.* Washington, DC: Bureau of Social Science Research.

Cantor., D., O'Hare, B. and O'Connor, K. (2008) "The Use of Monetary Incentives to Reduce Non-Response in Random Digit Dial Telephone Surveys" pp. 471-498 in J. M. Lepkowski, C. Tucker, J. M. Brick De Leeuw, E., Japec, L., Lavrakas, P. J., Link, M. W., & Sangster, R. L. (Eds.), *Advances In Telephone Survey Methodology,* New York: J.W. Wiley and Sons, Inc

Clark, Herbert H., and Susan E. Brennan. (1991) Grounding in communication. Perspectives on Socially Shared Cognition. Contributors: Lauren B. Resnick - editor, John M. Levine - editor, Stephanie D. Teasley - editor. American Psychological Association. Washington, DC.

Dillman, D. and J. Tarnai (1991) "Mode Effects of Cognitively Designed Recall Questions: A Comparison of Answers to Telephone and Mail Surveys." In Biemer, P., et al, Measurement Errors in Surveys. New York: John Wiley pp. 73 – 93.

Mingay, D. J. (2000). Is telephone audio computer-assisted self interviewing (T-ACASI) a method whose time has come? Paper presented at the 55th Annual Conference of the American Association for Public Opinion Research & World Association for Public Opinion Research, Portland, Oregon.

Montaquila, Jill, M., J. Michael Brick, Douglas Williams, Kwang Kim, and Daifeng Han. (2013) "A Study of Two-Phase Mail Survey Data Collection Methods." Journal of Survey Statistics and Methodology. 1(1), 66-87.

Schneider, S.J., Cantor, D. and T. Hagerty-Heller (2005). "Interactive Voice Response" pp. 349-356 in Polling America: An Encyclopedia of Public Opinion (Eds. Radcliff, B., Best, S.), Greenwood Press.

Singer, E., Van Hoewyk, J., and Maher, M. (2000). Experiments with Incentives in Telephone Surveys. Public Opinion Quarterly, 64:171-188.

Tourangeau, R., Steiger, D. M., and Wilson, D. (2002). Self-administered questions by telephone. Public Opinion Quarterly, 66, 265-278.

Westat, 2013. National Survey of Crime Victimization Companion Study – Pilot, Summary of Pilot Results, Report prepared for the Bureau of Justice Statistics, March 12, 2013.